

# 第1回 医療等情報の二次利用に関する技術作業班

## 日本の医療情報データ 二次利用の現状と課題

2024.2.8

東京大学情報基盤センター

データ科学研究部門 客員研究員

清水 央子



# リアルワールドデータ（RWD）と医療等情報データ

**リアルワールドデータ（RWD）**とは、日常の実臨床の中で得られる医療データの総称。レセプトデータや電子カルテのデータなど、様々な情報源から収集される患者さんの健康状態や医療行為のデータが含まれる。臨床試験とは異なり、複数の疾患や診断名のない症状も記録されており、医療ビッグデータとして活用が期待されている  
リアルワールドデータの例としては、以下のようなものがある

- レセプトデータ
- 電子カルテのデータ
- 患者レジストリデータ
- 健診データ
- ウェラブルデバイスから得られるバイタルデータ

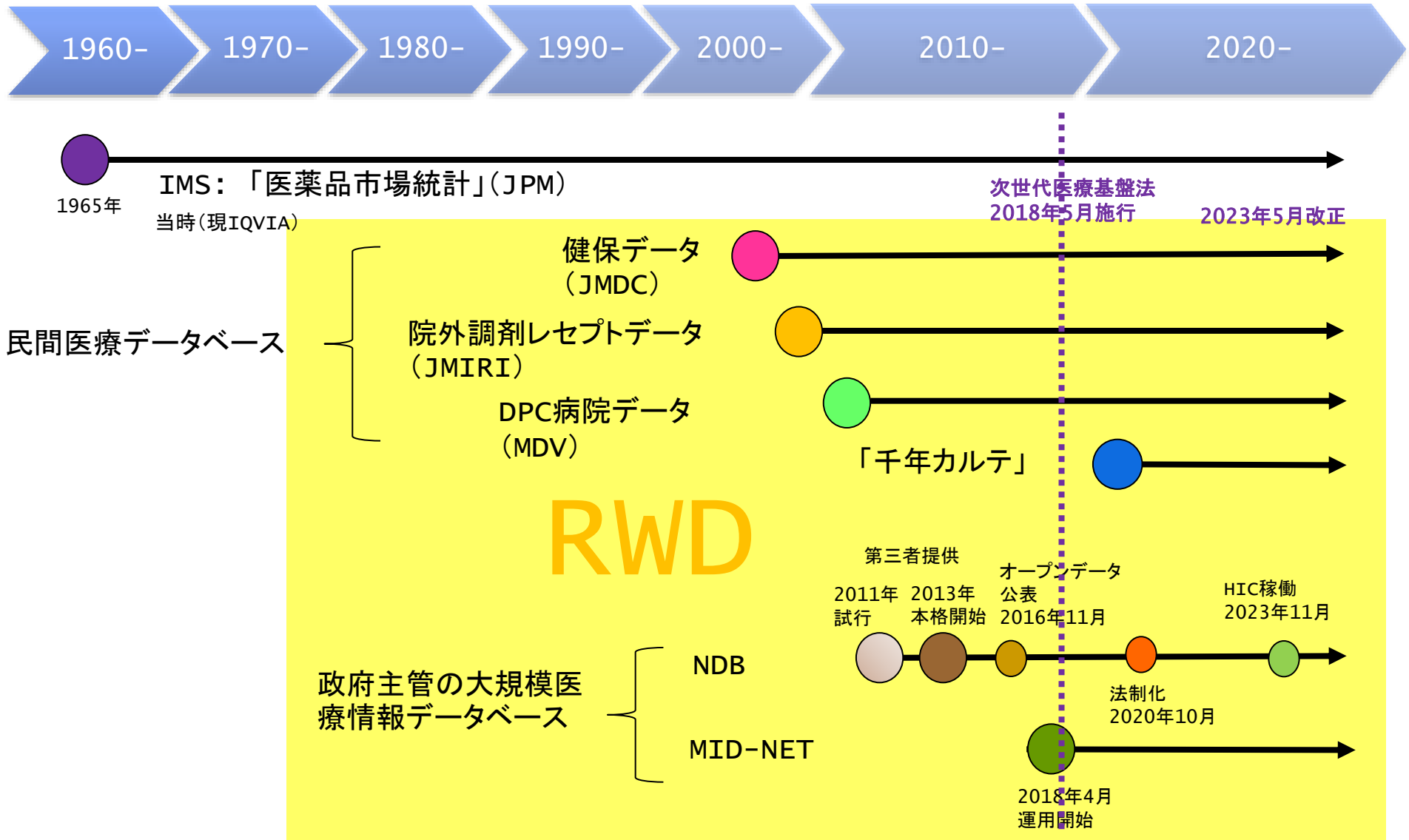
※MS Bingが「リアルワールドデータとは」で作成した資料（清水が一部改訂）

※逆に「臨床試験データ」は、実臨床の場ではなくコントロールされている下での治療結果なのでRWDには含まれない

👉 今回のWGでの議論の対象は「RWD（他医療関連情報）」の「二次利用」



# 日本における医療情報データの歴史





# 「日本における臨床疫学・薬剤疫学に応用可能なデータベース」

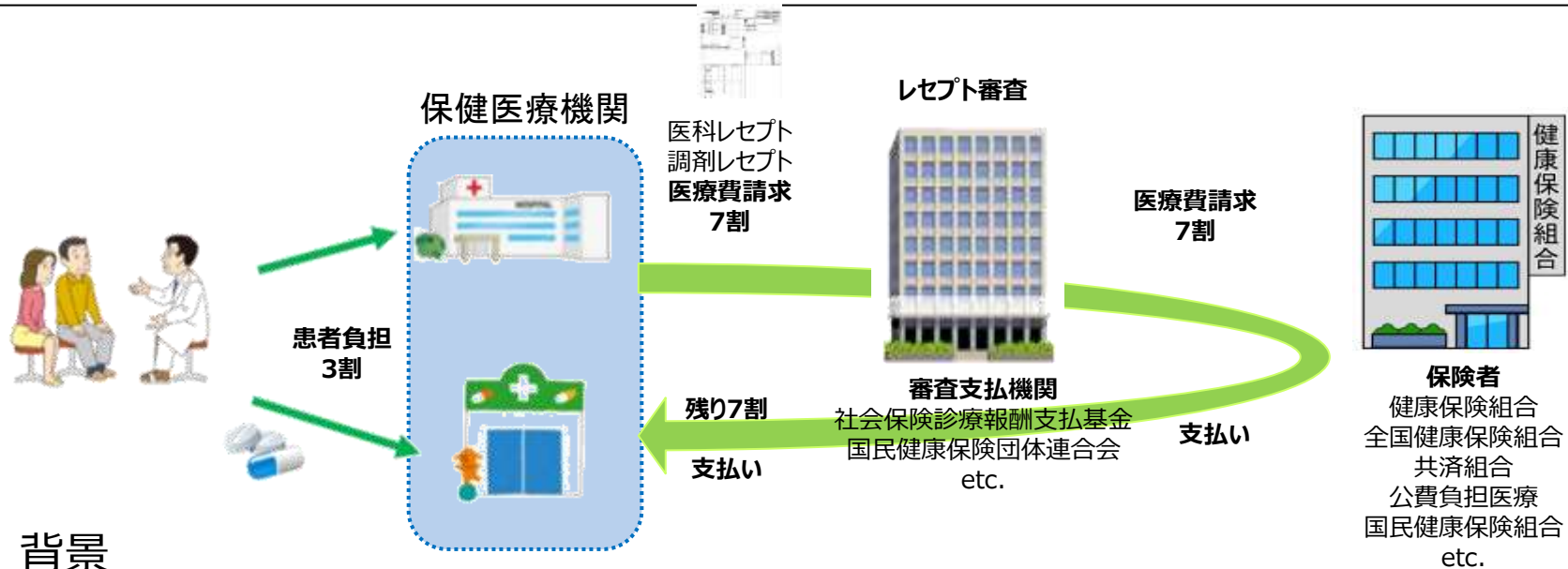
日本薬剤疫学会ホームページに掲載されたデータベース(他のリンクにないものはすべて日本薬剤疫学会ホームページに掲載されている)

	医療機関データベース										保健者データベース										その他	
	NHO NCD/MIA	4DR	HCD/RWD	LDI	MID-NET	値観値	MDV	JMDC	NCD	NDB	JMDC	Cross Fact	MDV	DeIC	IQVIA Claims data	Cross Fact	IQVIA NPA data	日本製薬 総合研究所	JMDC	HPWS	PatentsMap	
管理者(組織や会社など)の名称	国立病院機構	4DR	リアルワールドデータ株式会社	一般社団法人ライフデザインシステムズ	医療法人社団 医療情報総合機構	医療法人社団 医療情報総合機構	メディカル・データ・ビジョン株式会社	株式会社JMDC	一般社団法人 National Clinical Database	厚生労働省	株式会社JMDC	株式会社インテリリアルワールド	JamNet	メディカル・データ・ビジョン株式会社	DeIC	IQVIA ソリューションズ ジャパン株式会社	株式会社インテリリアルワールド	株式会社日本製薬総合研究所	株式会社JMDC	5SRH Center Envis	株式会社 社会情報サービス	
データベース名称	NHO NCD/MIA	4DR Research Network	RWDデータベース	千年カルテ	Medical Information Database Network (MED-IN)	医療法人社団 医療情報総合機構	EMM Prossard	JMDC医療情報データベース	National Clinical Database	レセプト情報・特定保健指導データベース	JMDC医療者データベース	Cross Fact	JamNetデータベース	EMM Prossard	DeIC Database	IQVIA Claims data	Cross Fact	IQVIA NPA data	JMDC医療者データベース	HPWS (National Health and Survey Database)	PatentsMap	
連絡先(代表メールアドレス等)	730-dip@ncd.mhl.go.jp	info@4dr.com	info@rwd.com	contact@ldi.or.jp	info@mid-net.or.jp	info@mid-net.or.jp	em_prossard@emm.co.jp	info@jmdc.co.jp	info@ncdb.or.jp	www.ncdb.or.jp	info@jmdc.co.jp	www.intelreal.com	support@jam-net.co.jp	em_prossard@emm.co.jp	www.deic.or.jp	www.iqvia.com	www.intelreal.com	info@jmdc.co.jp	www.hpws.com	patentsmap@patentsmap.com		
ウェブサイトURL	http://www.ncd.mhl.go.jp/ncd/000001.htm	http://www.4dr.com	http://www.rwd.com	http://www.ldi.or.jp	http://www.mid-net.or.jp	http://www.mid-net.or.jp	http://www.emm.co.jp	http://www.jmdc.co.jp	http://www.ncdb.or.jp	http://www.ncdb.or.jp	http://www.jmdc.co.jp	http://www.intelreal.com	http://www.jam-net.co.jp	http://www.deic.or.jp	http://www.iqvia.com	http://www.intelreal.com	http://www.jmdc.co.jp	http://www.hpws.com	http://www.patentsmap.com			
紹介論文DOI	DOI: 10.2147/CLIP.S339072								DOI: 10.1007/s10102014-423-4337													
データベースの概要	国立病院機構の施設の子カルテ情報と全施設の子カルテデータとを統合したデータベース	医療機関内の臨床データを集約、匿名化し、活用することを可能にするネットワーク	電子診療録、OPC、レセプト等の医療情報を統合したデータベース	特定保健指導の電子診療録(EMM)が、医療法人社団 医療情報総合機構(EMM Prossard)が、レセプトデータをソースとするデータベース	独立行政法人医薬品医療総合機構(AMED)が、医療法人社団 医療情報総合機構(EMM Prossard)が、レセプトデータをソースとするデータベース	特定保健指導の電子診療録(EMM)が、医療法人社団 医療情報総合機構(EMM Prossard)が、レセプトデータをソースとするデータベース	特定保健指導の電子診療録(EMM)が、医療法人社団 医療情報総合機構(EMM Prossard)が、レセプトデータをソースとするデータベース	特定保健指導の電子診療録(EMM)が、医療法人社団 医療情報総合機構(EMM Prossard)が、レセプトデータをソースとするデータベース	特定保健指導の電子診療録(EMM)が、医療法人社団 医療情報総合機構(EMM Prossard)が、レセプトデータをソースとするデータベース	特定保健指導の電子診療録(EMM)が、医療法人社団 医療情報総合機構(EMM Prossard)が、レセプトデータをソースとするデータベース	特定保健指導の電子診療録(EMM)が、医療法人社団 医療情報総合機構(EMM Prossard)が、レセプトデータをソースとするデータベース	特定保健指導の電子診療録(EMM)が、医療法人社団 医療情報総合機構(EMM Prossard)が、レセプトデータをソースとするデータベース	特定保健指導の電子診療録(EMM)が、医療法人社団 医療情報総合機構(EMM Prossard)が、レセプトデータをソースとするデータベース	特定保健指導の電子診療録(EMM)が、医療法人社団 医療情報総合機構(EMM Prossard)が、レセプトデータをソースとするデータベース	特定保健指導の電子診療録(EMM)が、医療法人社団 医療情報総合機構(EMM Prossard)が、レセプトデータをソースとするデータベース	特定保健指導の電子診療録(EMM)が、医療法人社団 医療情報総合機構(EMM Prossard)が、レセプトデータをソースとするデータベース	特定保健指導の電子診療録(EMM)が、医療法人社団 医療情報総合機構(EMM Prossard)が、レセプトデータをソースとするデータベース	特定保健指導の電子診療録(EMM)が、医療法人社団 医療情報総合機構(EMM Prossard)が、レセプトデータをソースとするデータベース	特定保健指導の電子診療録(EMM)が、医療法人社団 医療情報総合機構(EMM Prossard)が、レセプトデータをソースとするデータベース	特定保健指導の電子診療録(EMM)が、医療法人社団 医療情報総合機構(EMM Prossard)が、レセプトデータをソースとするデータベース	特定保健指導の電子診療録(EMM)が、医療法人社団 医療情報総合機構(EMM Prossard)が、レセプトデータをソースとするデータベース	
データベースの規模	約1000施設、約1000万人	約1000施設、約1000万人	約1000施設、約1000万人	約1000施設、約1000万人	約1000施設、約1000万人	約1000施設、約1000万人	約1000施設、約1000万人	約1000施設、約1000万人	約1000施設、約1000万人	約1000施設、約1000万人	約1000施設、約1000万人	約1000施設、約1000万人	約1000施設、約1000万人	約1000施設、約1000万人	約1000施設、約1000万人	約1000施設、約1000万人	約1000施設、約1000万人	約1000施設、約1000万人	約1000施設、約1000万人	約1000施設、約1000万人	約1000施設、約1000万人	
データベースの更新	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	
データベースの更新頻度	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	
データベースの更新頻度	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	リアルタイム	

次々と新しいデータベースが構築されています



# 日本の医療情報の中心は「レセプト」



## 背景

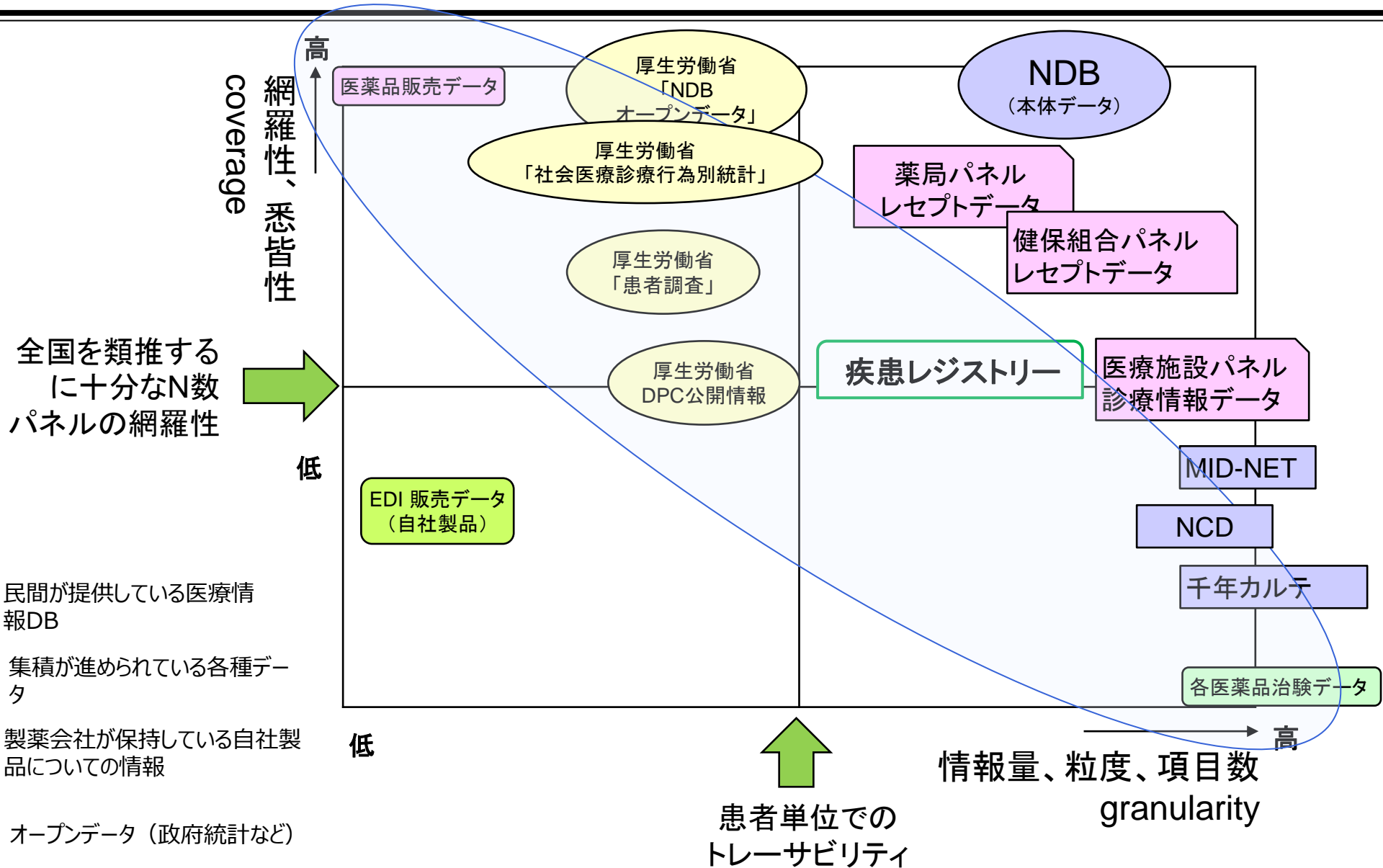
- 国民皆保険制度の下、全国民の保険医療についての情報がレセプトに反映されている → 複数の医療機関における治療履歴が追える
- 医療費の患者負担 3 割の残り 7 割の「請求書」なので、最低限必要な情報が記載されている → データが“きれい” (vs カルテは自由記載)
- (カルテ等に比べて) 電子化が進んでいる

## ただし

- レセプト病名
- 治療内容はわかるが、そのアウトカムや転帰については明快でない場合も  
例：検査をしたことはわかるが結果は記載されていない
- 保険組合を移動したら追うことが難しい → 長期にわたる治療記録が取れないことも



# 日本の医療情報データ相関図





# 日本のRWDの中心は民間のデータ

## 民間の医療情報データは

※データベース、ベンダーにより差がありますので、詳細は各社にお問い合わせください

- データが「きれい」
  - 集積したデータのクレンジングがされている
  - 主なマスターデータを裏に有している
- 入手までの期間が短く、データの速報性が高い
- (データを良く知る) ベンダー側に解析まで依頼することも可能  
(※ただし追加料金必要)

一方で

- 代表性をよく理解しないと落とし穴も...
- とても高価... 



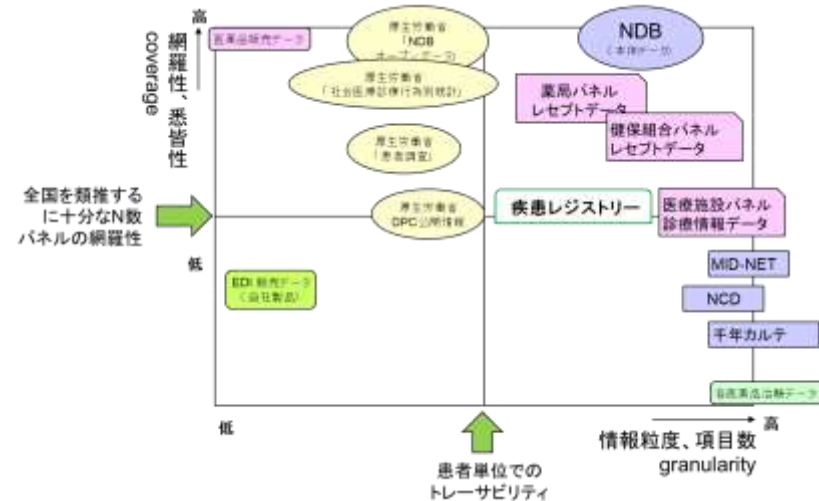
👉 「高かろう良かろう」の民間データ (vs 「安かろう悪かろう」)



# 日本の『医療』の“ビッグ”データは存在するのか？

## “スモール”データが散在し、トータルで“ビッグ”データとなっている状況

- ICTの進歩もあり、各種の医療や健康に関する大量な情報を、容易にタイムリーに補足し集積することが可能となってきた
- しかしながら個々のDBから得られる情報は限定的



ただし、これらのRWDをうまく組み合わせれば、情報がリッチになり、真の“ビッグ”データとなり得る

→ IoTの活用によりさらに有益なデータが集積され、またAIを駆使した解析により有益な知見が得られる可能性も👉



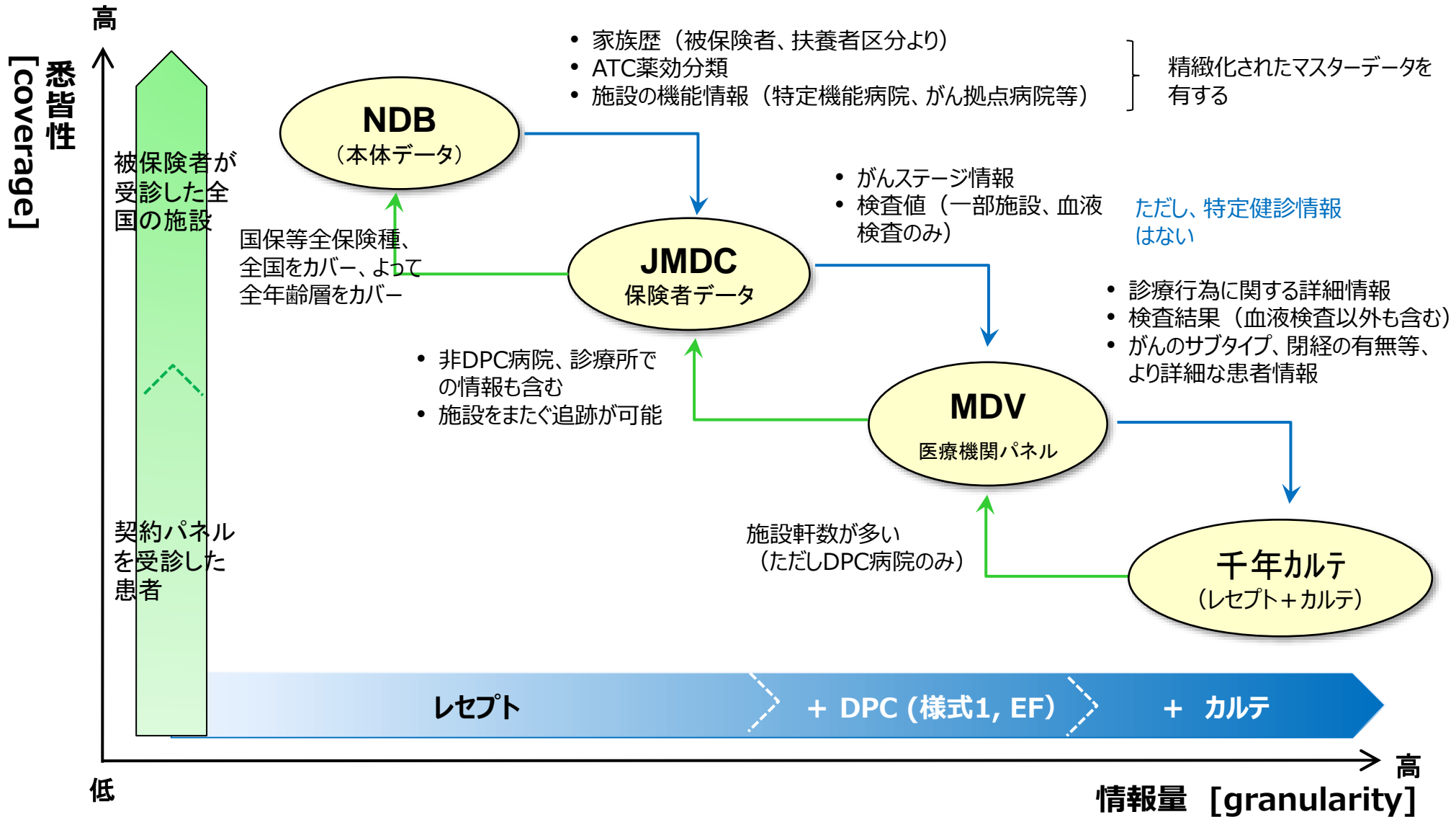


# レセプトデータの比較 ~NDB vs JMDC/MDV~

MDV		NDB		JMDC
約450施設、年間約1100万人の診療記録	<<	N数は国内最大(保険医療のほぼ全数が対象)	>>	約850万人の組合員のレセプトデータ
データ集積期間:2010.4-	≐	データ集積期間:2009.4-	<	データ集積期間:2005.1-
すべての保険組合が対象 保険外も一部カバー	>	すべての保険組合が対象	>>	組合健保のみ → 高齢者のデータが少ない
患者IDは施設毎に精緻化されている	>	患者ID1+ID2の突合プロセスが必要	<<	患者IDは保険組合毎に精緻化されている
当該施設への転院前・後は追えない	<<	長期の履歴が追える	>	保険組合の転入前、転出後は追えない
マスターデータが整備されている	>>	マスターデータは解析者が作成	<<	マスターデータが整備されている
高価 ※解析まで依頼することも可能	<<	安価	>>	高価 ※解析まで依頼することも可能
契約すれば数週間程度でデータの取得が可能	>>	申請後データ取得までの1年以上かかる	<<	契約すれば数週間程度でデータの取得が可能
直近データは3-4か月遅れ	>>	直近データは5-6か月遅れ。ただし申請内容やタイミングによりもっと遅れることも	<	直近データは5-6か月遅れ
個人単位	=	個人単位	<	家族(被保険者-扶養者)関係がわかる



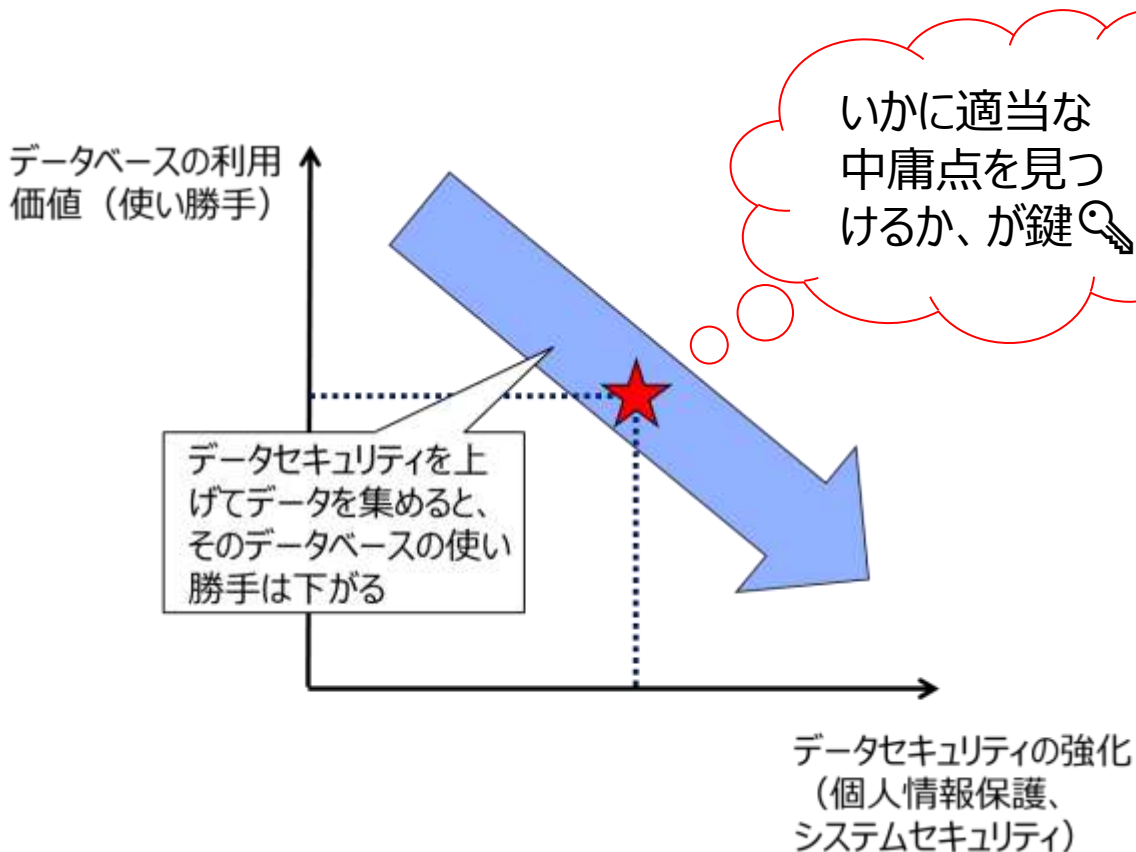
# 「カルテ」データ含めた各種RWDの優位性比較





# セキュリティの強化 vs 使い勝手

データを集積（DB化）すれば、その利用が進むわけではない！  
一方で、より質の良い（粒度・悉皆性の高い）データを提供するために、セキュリティ要件を強化しすぎ使い勝手が著しく低下しては利用の推進にはつながりにくい



## 👉 利用推進のための条件

- ①データ自体の質の向上
- ②適正なセキュリティ要件に基づく利用環境の整備（データ提供体制、解析環境）



# ①データ自体の質の向上

「データを集める」だけでは利用は進まない

利活用者から見たデータの標準化、Visiting環境での利用環境のあり方等について

課題：データ自体の質の向上

- 粒度 and/or 悉皆性を上げる↑  
複数DBの連結（例：NDB+介護）→ そのためにはデータの標準化が必要
- マスターの整備（例：NDBの診療報酬/薬剤/医療機関マスター等、JLAC10）
- 情報の質（例：検査結果@カルテ）→ インput情報（データ入力時）の質の向上 → 医療現場からの協力も必要なのでは？（vs 現場の負担増）
- 精緻化された患者IDによる同一患者の特定と連結

👉 現時点では、利用者側にとって本来の解析の前の下準備の段階での負荷が大きく、誤った解析をしてしまう可能性も



# NDB解析に必要な「マスターデータ」

例示

主要マスターデータ	一般公開されているマスターの種類
傷病名マスター (約26,000レセ電処理用傷病名)	「診療報酬情報提供サービス(厚労省保険局)」1) 「社会保険診療報酬支払基金」1) 2) 「医療情報システム開発センター[MEDIS]」2) 3)
医薬品マスター (約21,000の医療用医薬品)	
診療報酬マスター (医科診療行為: 約7,500)	
医療機関マスター (全国約18万医療施設)	全国版のマスターは存在しない 研究者各自が各地方厚生局のサイトより47都道府県分の医療機関リストをダウンロードし統合

共通課題: いずれもある時点でのマスターであり  
過去に遡った累積情報を収載していない

データ解析という“二次利用”を想定したマスターデータが整備されていない

- 1) 別サイトで運用されているがソースは同じ
- 2) 「傷病名」においては共同で更新
- 3) 「診療報酬」においては、手術・検査など一部の項目のみ



# 「グランドマスター」の必要性

例示

RQ「DPC病院とそれ以外で〇〇の治療方法の差異を調べたい」

RQ「抗インフルエンザ薬の各シーズンの投与患者数の推移を知りたい」

『医療機関マスター』によるフラグ付け

『医薬品マスター』より過去のコードも含めて抽出



1. 地方厚生局URLより各都道府県の医療施設のリストと付随情報をD/LLし、
2. DPC病院のリストと照合してリストを作成

\* 解析対象期間中に存在したすべてのコードを抽出することが必要

年度	都道府県番号	点数表	医療機関番号	医療機関名	DPC対象施設フラグ
2020	01	1	1234567	〇〇〇大学病院	1
2020	01	1	4567890	△△△	
2020	02	1	3456789	●●●	
2020	02	3	7891234	▲▲▲	

年度	都道府県番号	点数表	医療機関番号	医療機関名	DPC対象施設フラグ
2019	01	1	5678901	〇〇〇大学病院	1
2019	01	1	4567890	△△△クリニック	0
2019	02	1	3456789	●●●市立病院	1
2019	02	3	7891234	▲▲▲デンタルクリニック	0

医療機関コードが変更  
※移転、経営母体の変更等に伴う

医薬品コード (レセプトコード)	医薬品名	後発品フラグ
660443018	リレンザ	0
610443074	タミフルカプセル	
610462002	タミフルドライシロ	
622012101	イナビル吸入粉	
622622501	ゾフルーザ錠	
622622601	ゾフルーザ	
621972202	ラビアクタ点滴静注液バイアル150mg	0
621972102	ラビアクタ点滴静注液バッグ300mg	0
621972101	ラビアクタ点滴用バッグ300mg	0
621972201	ラビアクタ点滴用バイアル150mg	0
622638801	オセルタミビルカプセル75mg「サワイ」	1
622638901	オセルタミビルDS3%「サワイ」	1

2013.3.31付で廃止された古いコード  
※販売会社、名称変更等に  
伴う



# 「カルテ」データは万能か？

例示

例：「カルテ」で検査結果が判明できたのは75%



## HR/HER2サブタイプの特定



サブタイプ「HR/HER2」の特定ができた割合は、A,B,Cの3施設の中で差が見られた。また当該施設で検査を行った場合は8割以上でサブタイプ「HR/HER2」の判定ができたが、当該施設で検査がなかった場合は半分以下となっており、大きな差が見られた。

	対象患者	HR/HER2 判定可能患者数*		当該施設での 検査あり**	検査なし	合計
A病院	113	93 (82%)	HR/HER2 判定可能*	627 (81%)	72 (46%)	699 (75%)
B病院	421	320 (76%)	HR/HER2 判定不可	147 (19%)	84 (54%)	231 (25%)
C病院	396	286 (72%)				
合計	930	699 (75%)	合計	774 (100%)	156 (100%)	930 (100%)

\* HR/HER2両者が判定可能なケースのみ

\*\* レセプトで確認



## ②解析環境の整備

セキュリティ要件が厳しすぎると使い勝手が悪くなり、利用が進まない

🏠 情報の価値を最大限保ちつつセキュリティ要件を満たすための“落としどころ”を探る必要

「セキュリティ」管理の対象：

- A) 悪意ある第三者による情報の不正取得
  - ✓ システムに侵入（ハッカー）
  - ✓ 管理部屋に物理的な侵入
  - ✓ タコバヤでの窃視、聞き耳による → これで情報が盗めるのか？
- B) 悪意ある利用者による目的外使用 → 出口管理で対応（案）
- C) データから個人が特定できてしまう可能性
  - ✓ 希少疾病（N数が少ない）
  - ✓ 最小集計単位の設定
  - ✓ それでも特定できることはある（専門医の専門分野の場合等）

- 申請手続きの簡素化と審査の迅速化  
（例：NDBでは申請からデータ提供までのリードタイム1年以上、申請は研究機関単位＝学長が申請者という体）
- タコバヤの過度なセキュリティ要件の見直し
  - セキュアなネットワークとは？
  - タコバヤの規制
    - ✓ 入退室管理はどこまで必要か
    - ✓ 窃視、聞き耳対策は必要か？
- 個人が特定できてしまう可能性を排除するための規制
  - 最小集計単位の原則（K匿名化）
  - 罰則はどこまで？
  - 特定できる可能性があってはNG、では解析の幅が広がらない可能性も





# NDBの解析は「三密」の環境で行うことが原則

例示

三密？

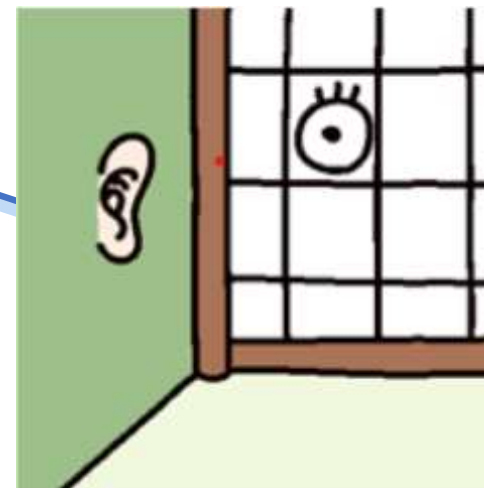
NDB専用室

解析用ソフトウェア

SQL、SAS、Python、R、JMP、  
SPSS... も必要



バックアップ用HDDも必要



©ことわざ・慣用句の百科事典

※外部とネットワークでつながって  
てはいけない → ソフトウェアのア  
プデートもままならない

リモートワークでは解析作業は行うことができない...



# ご参考：当研究室のNDBでの経験

例示

セキュアなネットワークでのアクセスが不承諾となった例：

申請したシステム構成図（この他にセキュリティに関する説明を加えて提出）



I型認定に近いシステム構成

## 不承諾の理由：

プライベートクラウドの利活用を必ずしも否定するものではないが、**取扱者自らが設置したものではない**うえ、設置主体、保守運営事業者、情報基盤センターおよび利用者の**責任分解点が明らか**ではなく、セキュリティインシデント等の**不備の事態が発生した場合における体制や対応についても不明な点が多数存在**することを指摘せざるを得ない。

このような状況下において「匿名レセプト情報・匿名特定健診等情報の提供に関するガイドライン」で定めるセキュリティ要件を満たしていると認められないため、本件申し出について不承諾が相当と判断したもの。

厚生労働省発保0705第2号より



不承諾の理由について納得がいかなかったが、以下の条文により**不服申し立ても断念**

匿名レセプト情報等の提供は厚生労働大臣と提供申出者（利用者）及び取扱者の双方の合意の基づく契約上の行為であり、行政手続法（平成5年法律第88号）第2条第2号の処分にならないことから、**行政不服審査法（昭和37年法律第160号）の対象外**となりますことご了承ください



# 「K匿名化」の例

例示

## 「NDBオープンデータ」 医薬品の使用量（例）

1000単位（錠剤の場合は「錠数」）未満をK匿名化（最小集計単位の原則）

→ 「980か355か」がわかるとどうやって個人の特定ができるのか？

薬効分類	薬効分類名称	医薬品コード	医薬品名	単位	薬価基準収載 医薬品コード	薬価	後発品 区分	総計	01			02			03				
									県別合計	北海道	青森県	岩手	岩手	岩手	岩手				
235	下剤, 浣腸剤	620480601	センノシド錠 1.2mg 「サワイ」	錠	2354003F2367	5.0	1	228,215,579	228,215,579	10,984,513	4,691,918	3,5							
		620481301	ブルゼニド錠 1.2mg	錠	2354003F2316	5.6	0	144,253,041	144,253,041	7,739,994	2,500,621	1,5							
		622182901	アミティーザカプセル 2.4μg	カプセル	2359006M1025	123.0	0	140,085,864	140,085,864	9,045,733	1,937,566	1,6							
		620491801	アローゼン顆粒	g	2359102X1072	7.2	0	83,457,982	83,457,982	12,607,997	1,037,764	1,3							
		620006963	センナリド錠 1.2mg	錠	2354003F2251	5.0	1	78,667,155	78,667,155	3,411,686	1,202,166	1,1							
		620008638	センノサイド錠 1.2mg	錠	2354003F2278	5.0	1	64,176,581	64,176,581	4,291,427	1,200,831	4,							
		621278001	センノシド錠 1.2mg 「トーワ」	錠	2354003F2375	5.0	1	62,800,082	62,800,082	5,535,207	999,143	9							
		612350042	ヨーデルS糖衣錠-80 80mg	錠	2354002F2028	6.5	1	62,096,053	62,096,053	590,726	50,051	1							
		622564801	グーフイス錠 5mg	錠	2359008F1025	105.8	0	57,566,025	57,566,025	4,280,332	573,429	1,							
		620480205	センノシド錠 1.2mg 「セイコー」	錠	2354003F2340	5.0	1	55,339,486	55,339,486	4,348,619	1,686,154	1,							
		620480707	センノシド錠 1.2mg 「武田テバ」	錠	2354003F2448	5.0	1	44,620,761	44,620,761	2,719,457	865,331	3							
		620481202	センノシド錠 1.2mg 「ファイザー」	錠	2354003F2413	5.0	1	42,906,246	42,906,246	3,401,734	1,077,759								
		622625001	モビコール配合内用剤 6.8523g	包	2359110B1029	83.9	0	41,930,890	41,930,890	1,733,828	390,699								
		620006956	ヤチロ配合錠	錠	2359103X1034	5.6	0	36,670,635	36,670,635	2,964,931	65,170								
						錠	2354003F2332	5.0	1	34,995,144	34,995,144	2,820,024	1,402,261						
				622263901	センノシド錠 1.2mg 「クニヒロ」	錠	2356002X1099	1.2	0	17,013	17,013	15,467	-	-	-	-	-	-	-
				620006886	加香ヒマシ油「ケンコー」	mL	2359001X2017	7.2	0	17,013	17,013	15,467	-	-	-	-	-	-	-
		612350146	カルメロースナトリウム	g	235900501110	9.6	1	16,111	12,246	4,424	-	-	-	-	-	-	-		
		620488301	ビコスルファートナトリウム顆粒 1% 「イセイ」	g	2355002X1016	0.7	0	12,736	10,134	-	-	-	-	-	-	-	-		
		620481711	硫酸マグネシウム「東海」	g	2356001X1019	1.2	0	12,383	9,132	-	-	-	-	-	-	-	-		
		612350033	ヒマシ油	mL	2354003F2294	5.0	1	11,843	9,896	-	-	-	-	-	-	-	-		
		620480901	ソルダナ錠 1.2mg	錠	2359005F1170	5.8	1	9,818	9,526	-	-	-	-	-	-	-	-		
		620007051	フェニスルー錠 2.5mg	錠	2356001X1193	1.4	0	4,840	1,680	-	-	-	-	-	-	-	-		
		620006895	ヒマシ油（山善）	mL	235910981077	0.7	1	4,000	4,000	-	-	-	-	-	-	-	-		
		612350062	人工カルルス塩「コザカイ・M」	g	2356001X1124	1.4	0	2,320	1,700	-	-	-	-	-	-	-	-		
		620006883	ヒマシ油「東海」	mL	2356002X1013	1.3	0	1,610	1,000	-	-	-	-	-	-	-	-		
		612350048	加香ヒマシ油	mL	235910981115	0.7	1	1,270	0	-	-	-	-	-	-	-	-		
		612350065	人工カルルス塩「ヤマゼン」	g	2355002X1016	0.7	0	-	0	-	-	-	-	-	-	-	-		
		612350119	硫酸マグネシウム「カナダ」	g	2356002X1013	1.2	0	-	0	-	-	-	-	-	-	-	-		
		620482319	加香ヒマシ油「東海」	mL	2359005M1098	5.0	1	-	0	-	-	-	-	-	-	-	-		
		621278301	ビコスルファートナトリウムカプセル 2.5mg 「T.C」	カプセル	2356002X1188	1.7	0	-	0	-	-	-	-	-	-	-	-		
		620006893	加香ヒマシ油（山善）	mL	2356002X1013	1.3	0	-	0	-	-	-	-	-	-	-	-		
		620482310	加香ヒマシ油「ニコッ」	mL				-	0	-	-	-	-	-	-	-	-		



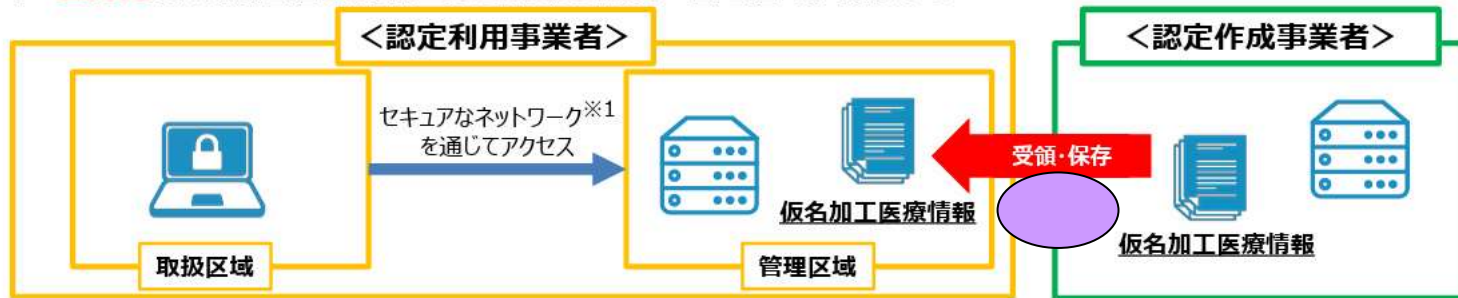
# 仮名加工の取り扱いを可能とする解析環境 次世代医療基盤法（内閣府）案

## IV. ③ 認定利用事業者の安全管理措置に関する基本的な考え方（1）

### ○ 「I型認定」「II型認定」及び「管理区域」「取扱区域」

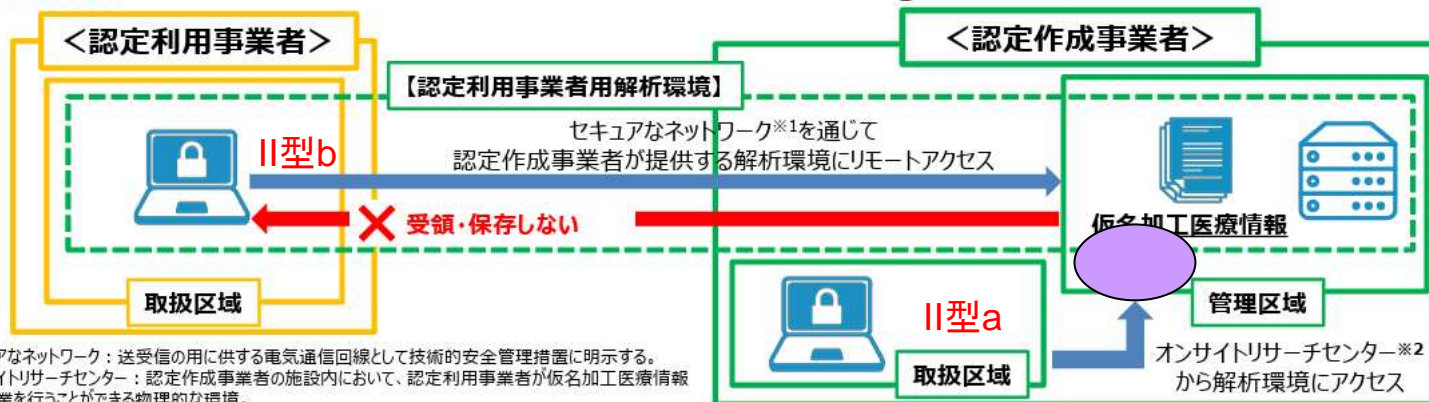
- ・ 確実な安全管理措置の確保と仮名加工医療情報の利活用促進の両立の観点から、利用事業者が自らの整備した環境下に仮名加工医療情報を保存することが可能な **I型認定**に加え、認定作成事業者等が整備したVisiting環境での利用に限定し、当該環境を前提とした安全管理措置の要件を定める **II型認定**の2種類を設ける（新GLIV-5-1-1-2参照）。
- ・ また、提供仮名加工医療情報※を取り扱う施設設備として、仮名加工医療情報を保存する機器が設置される**管理区域**と、提供仮名加工医療情報の操作・解析等を行う**取扱区域**を特定し、安全管理のために必要かつ適切な措置を講じることを求める（新GLIV-14-3-1-1参照）。 ※認定作成事業者から認定利用事業者へ提供された仮名加工医療情報

### 【I型認定】仮名加工医療情報を認定利用事業者の管理区域に保存する



個別の解析のために切り出したデータセット

### 【II型認定】認定利用事業者は仮名加工医療情報を保存せず、Visiting環境限定で利用



※1 セキュアなネットワーク：送受信の用に供する電気通信回線として技術的安全管理措置に明示する。  
 ※2 オンサイトリサーチセンター：認定作成事業者の施設内において、認定利用事業者が仮名加工医療情報を用いた作業を行うことができる物理的環境。

次世代医療基盤法検討WG資料より



# 解析環境準備の負荷を軽減するために

ご提案

前頁の仕組みにおける各事業者の負荷

	作成事業者	利用事業者
I型	データを切り出したらそれを利用者に渡す仕組み（アップロードあるいは郵送）までを整備	「管理区域」準備の負担最大
II型a	「取扱区域」（visiting作業現場）準備の負担最大	visitingの場所まで物理的に移動しないと作業ができない
II型b	リモートアクセスの受け側のシステム環境整備が必要	Thinkクライアントとセキュアなネットワーク環境の用意が必要

👉 解析環境の準備における両事業者の負担が大きすぎることにより、利用の推進を妨げる可能性も



共通の解析基盤を構築し、データ提供者、利用者の双方から利用料を徴収することにより運営する仕組みを整備

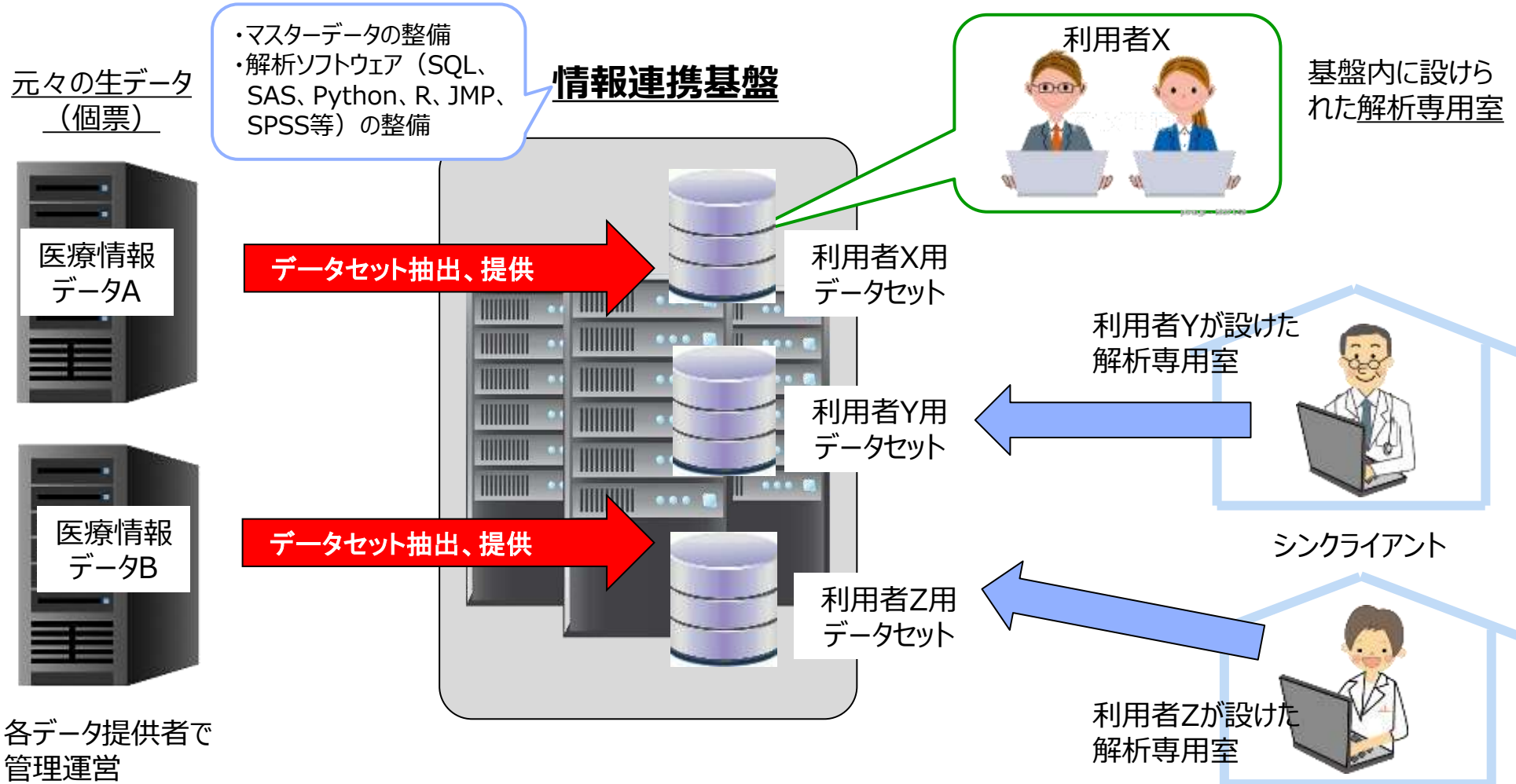
- ✓ 「セキュアなネットワーク」を定義し提供
  - ✓ 解析専用室はシンクライアントでデータを持たないため、窃視や聞き耳などの対策も不要
- 👉 結果として全体のコスト、負荷は軽減し、データ利用のハードルも下がる可能性



# 解析環境の整備：情報連携基盤（素案）

ご提案

解析環境の準備におけるデータ提供者、解析者双方の負担軽減のために**共通の解析基盤を整備**し、個別解析（利用）のために切り出したデータセットを格納。解析者はセキュアなネットワークでアクセスして個々に解析を行う





### 現在の稼働中サービス



Wisteria/BDEC-01

通常サービス中

#### システムの特徴

Wisteria/BDEC-01システムは、東京大学情報基盤センターが運用するスーパーコンピュータシステムです。Wisteria/BDEC-01はシミュレーションノード群（Odyssey（オデッセイ））とデータ・学習ノード群（Aquarius（アクエリアス））の2つの計算ノード群を有したシステムです。シミュレーションノード群（Odyssey）は、世界最高性能を有するスーパーコンピュータ「富岳」と同じ富士通株式会社の「FUJITSU Processor A64FX」を搭載した計算ノード7,680台から構成されます。データ・学習ノード群（Aquarius）はインテル社製第3世代Xeonスケーラブルプロセッサ（開発コード名Ice Lake）およびNVIDIA A100 Tensor コア GPUを搭載した計算ノード45台から構成されます。

東京大学情報基盤センター（以下、センター）では、現在Wisteria/BDEC-01 スーパーコンピュータシステムを運用しております。本スーパーコンピュータシステムでは、「通常利用（一般）」によるサービスを行っています。

🏠 大学・公共機関は、承認されれば安価（電気代程度）での利用が可能

データ提供者、利用者双方から相当の利用料を徴収し運営する仕組みを整備



**ご清聴ありがとうございました**

清水連絡先:

[eshimizu@g.ecc.u-tokyo.ac.jp](mailto:eshimizu@g.ecc.u-tokyo.ac.jp)

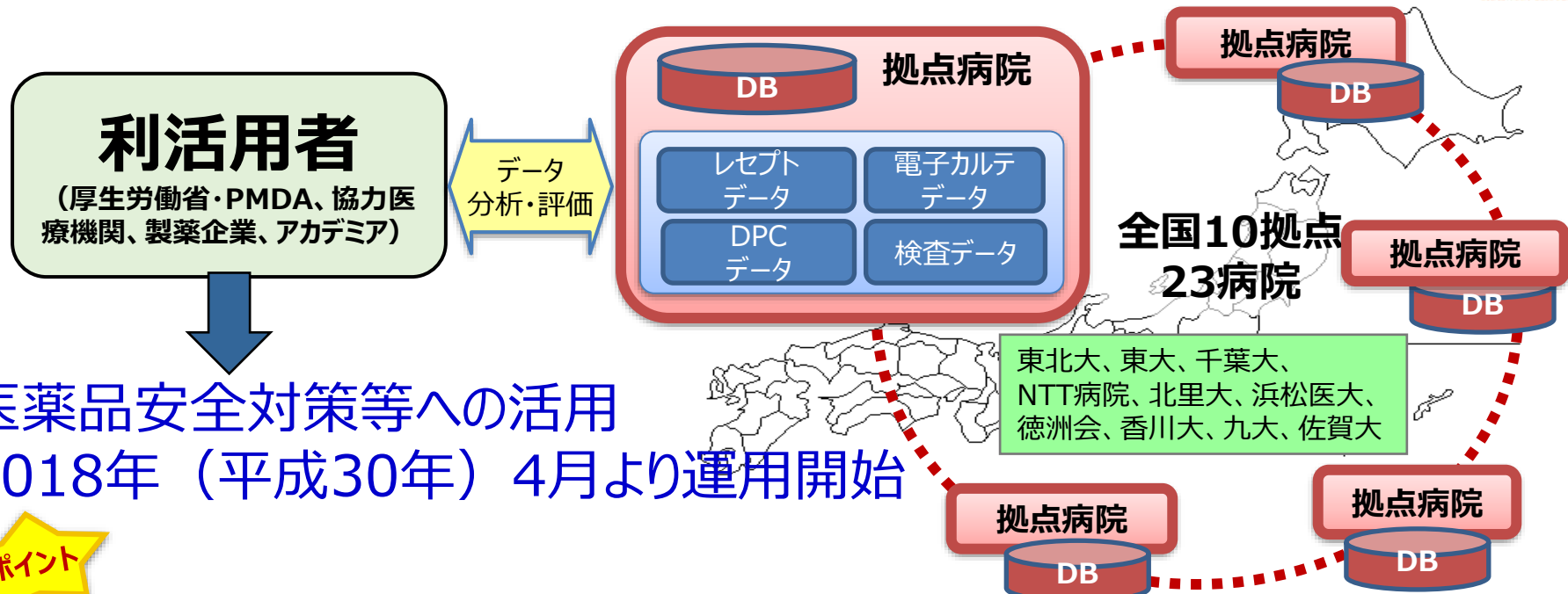
[eiko\\_shimizu@ds.itc.u-tokyo.ac.jp](mailto:eiko_shimizu@ds.itc.u-tokyo.ac.jp) (情報基盤センター内)



# MID-NET<sup>®</sup>の品質管理・標準化



独立行政法人医薬品医療機器総合機構（PMDA）  
医療情報科学部長 山口 光峰



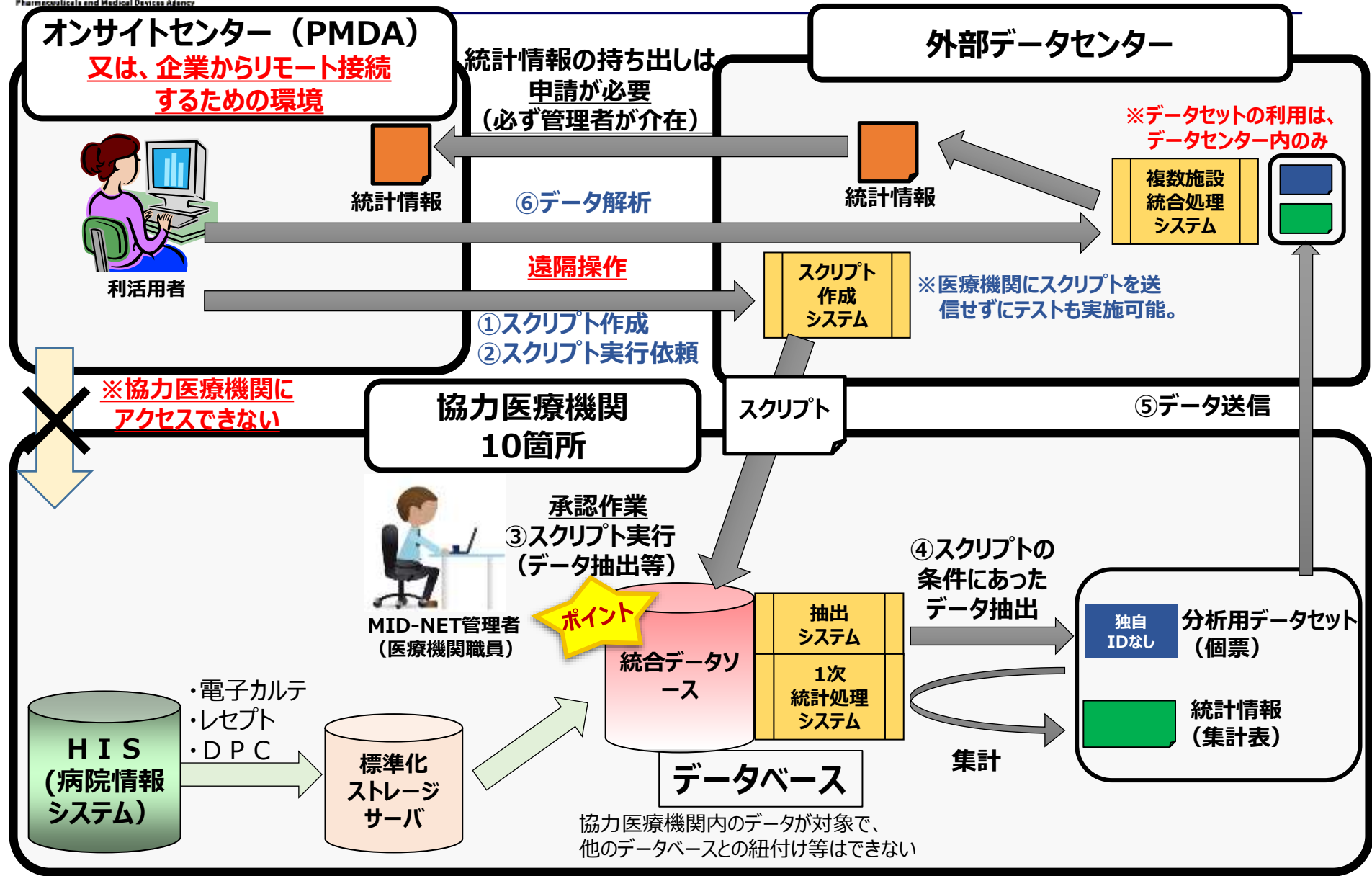
医薬品安全対策等への活用  
2018年（平成30年）4月より運用開始

**ポイント**

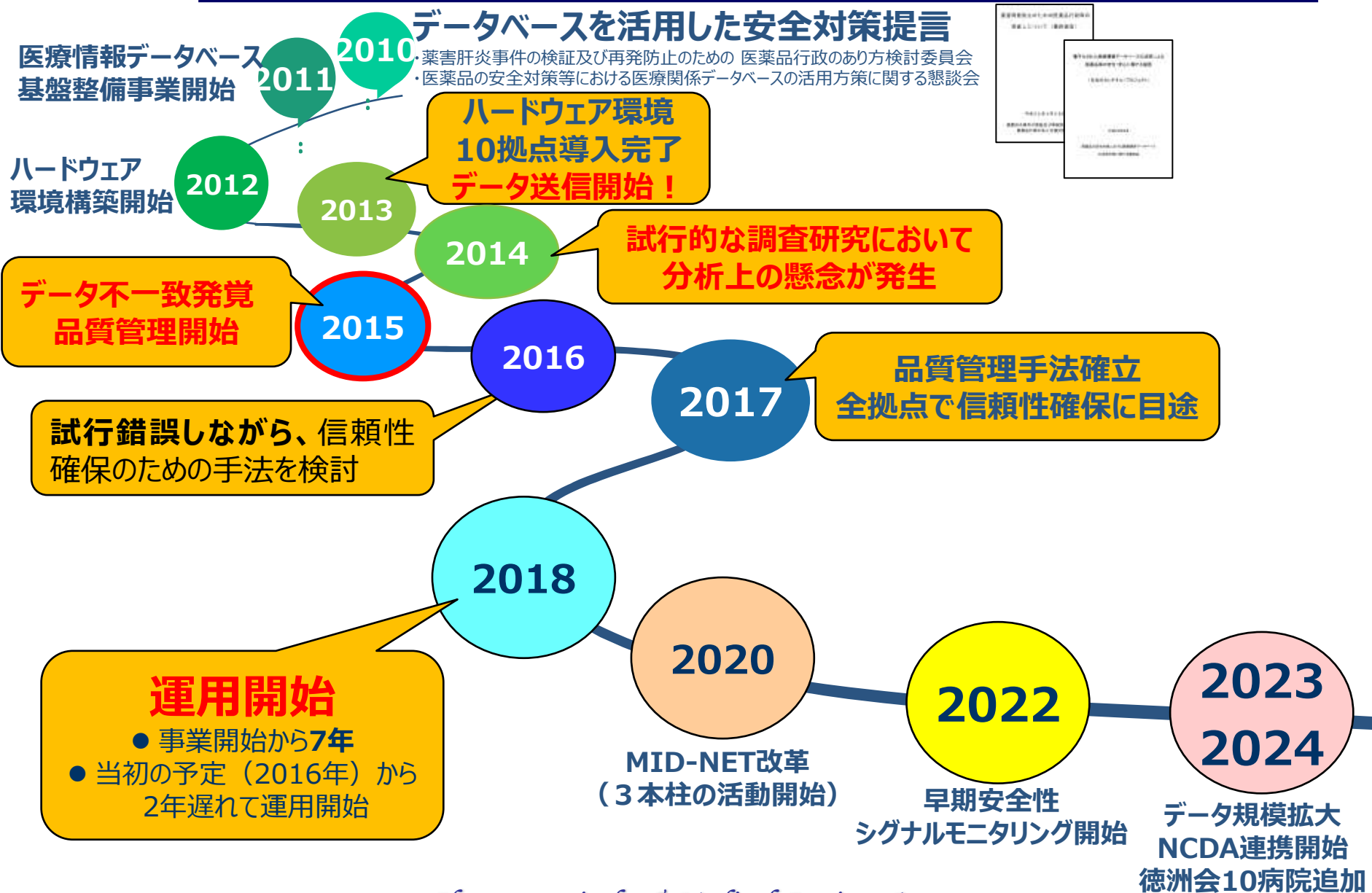
- PMDA法第15条第1項第5号ハ及びヘに基づく業務（※個人情報法の法令除外に基づく提供）
- 各拠点にデータベース（統合データソース）を設置
- 605万人超の規模（2022年12月末現在）→徳洲会10病院の拡大運用開始（2024年）
- **品質管理された医療情報データベース。薬機法の基準に基づき、高い信頼性を確保**
- **PMDA安全対策部が安全対策に活用中。また、製薬会社が製造販売後DB調査を実施中**
- **行政利活用において、定型的な解析の自動処理を運用中**
- 病名、処方等に加え、350項目以上の臨床検査結果が利用可能
- 専用ホームページにおいて、利活用等に必要な各種情報を掲載

<https://www.pmda.go.jp/safety/mid-net/0001.html>  
Pharmaceuticals & Medical Devices Agency

# MID-NETシステムの全体像



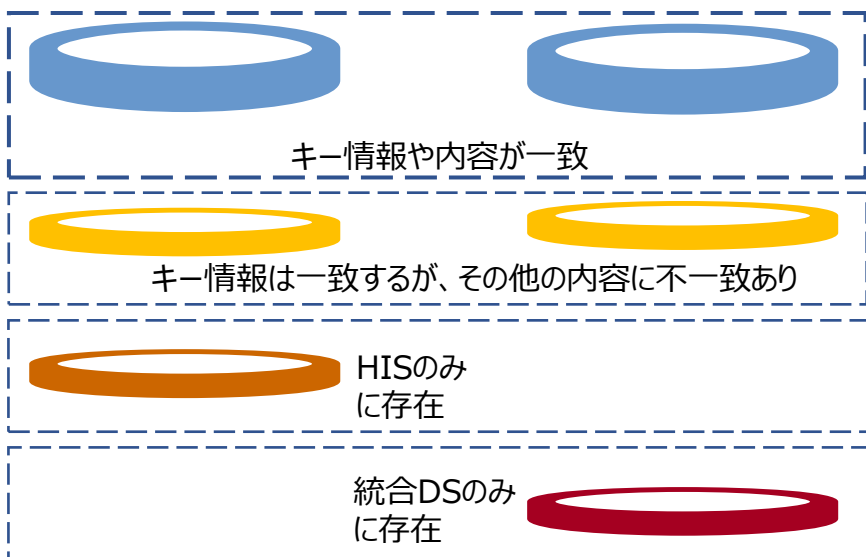
# MID-NET<sup>®</sup>運営の道程と品質管理



【2015年頃や電子カルテリプレース時】

## 品質管理前

病院情報システム (HIS) ↔ 比較 ↔ 統合データソース (統合DS)



## 品質管理後

病院情報システム (HIS) ↔ 比較 ↔ 統合データソース (統合DS)



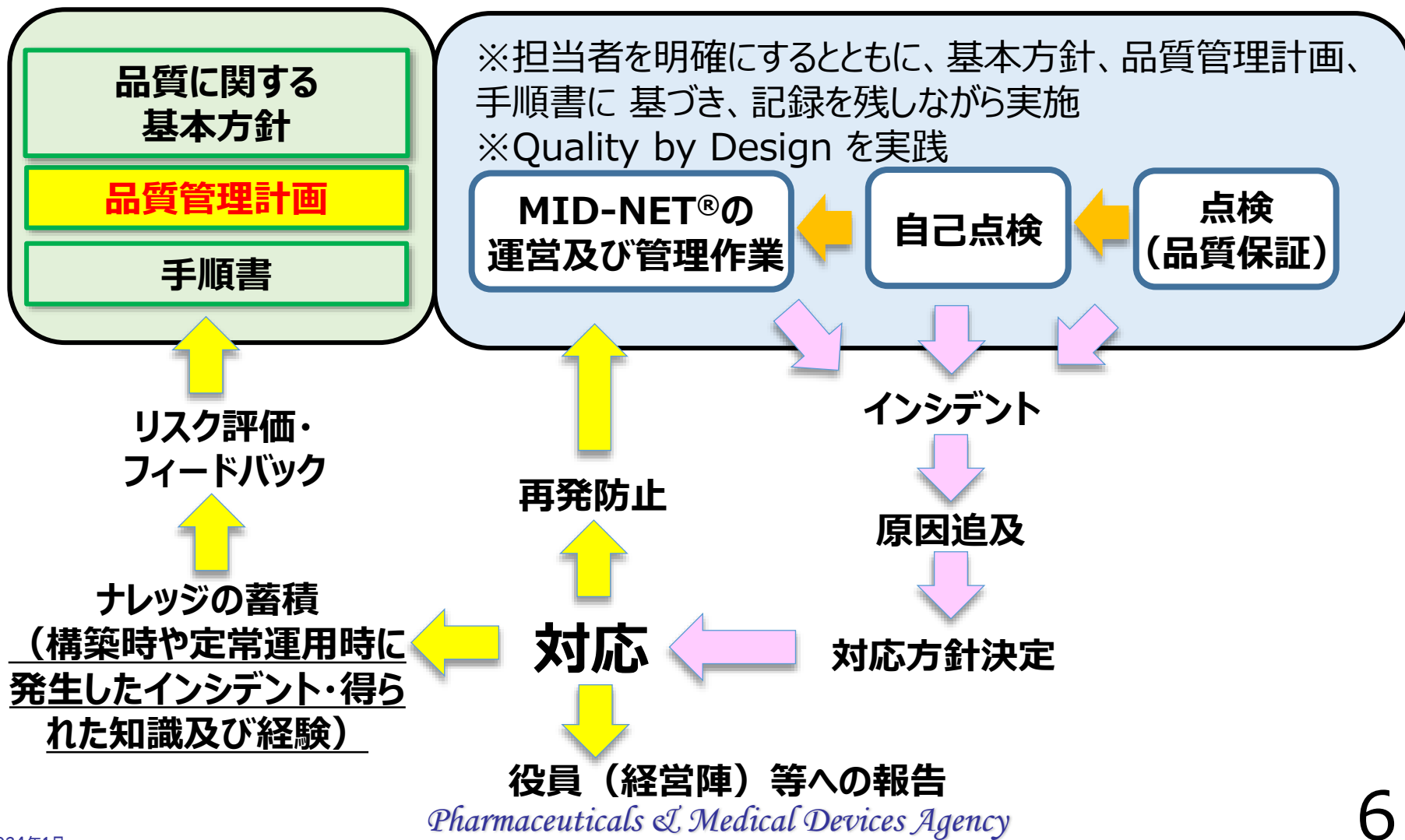
Real Worldでの状況に対応した品質管理が必要

- 元データとデータベースに集積されているデータでは、様々なパターンによる差異が認められた。

### 差異の例

- 処方中止のデータが通常の処方データと区別されずに送信されていた。
- 1日量、1回量、全量のデータの取り扱いが医療機関ごとに異なっており、データベースでは同じ項目に様々な用量が混在していた。
- 標準単位がないままデータが送信されていた。

MRDA (MID-NET<sup>®</sup> Real-time Data-quality Assurance)



## 1. システムの品質管理（原則、導入時・改修時）

データ送受信プロセスの各機能を検証

- ①データ取込機能 ②標準化機能 ③抽出機能 ④送信機能
- ⑤SAS変換機能

▶ ベンダーが実施する検証に加え、PMDAも別の角度から検証を実施

## 2. データの品質管理（オリジナルデータとの一貫性確認）

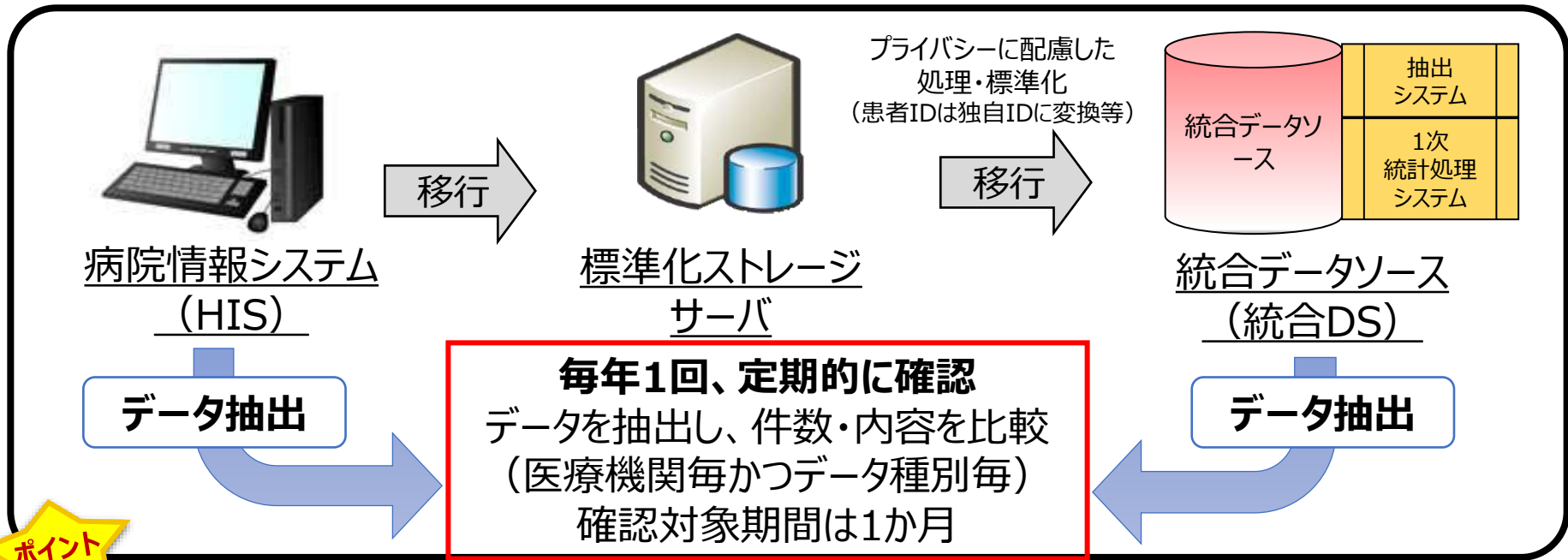
全データ種別（電子カルテ、レセプトおよびD P C）を対象に、定期的に元データとの一貫性を確認

▶ 電子カルテリプレース時には重点的に確認

## 3. 日常的品質管理（データとシステムの日常的モニタリング）

各プロセスの稼働状況を日々確認

▶ データ受信件数、データファイル格納状況、システム稼働状況、バックアップ状況、スクリプト処理状況、医療機関マスタ更新状況



※電子カルテのリプレイス、運用変更によりデータ送信状況は変化  
電子カルテデータは全医療機関で、レセプト・DPCデータは代表医療機関で確認

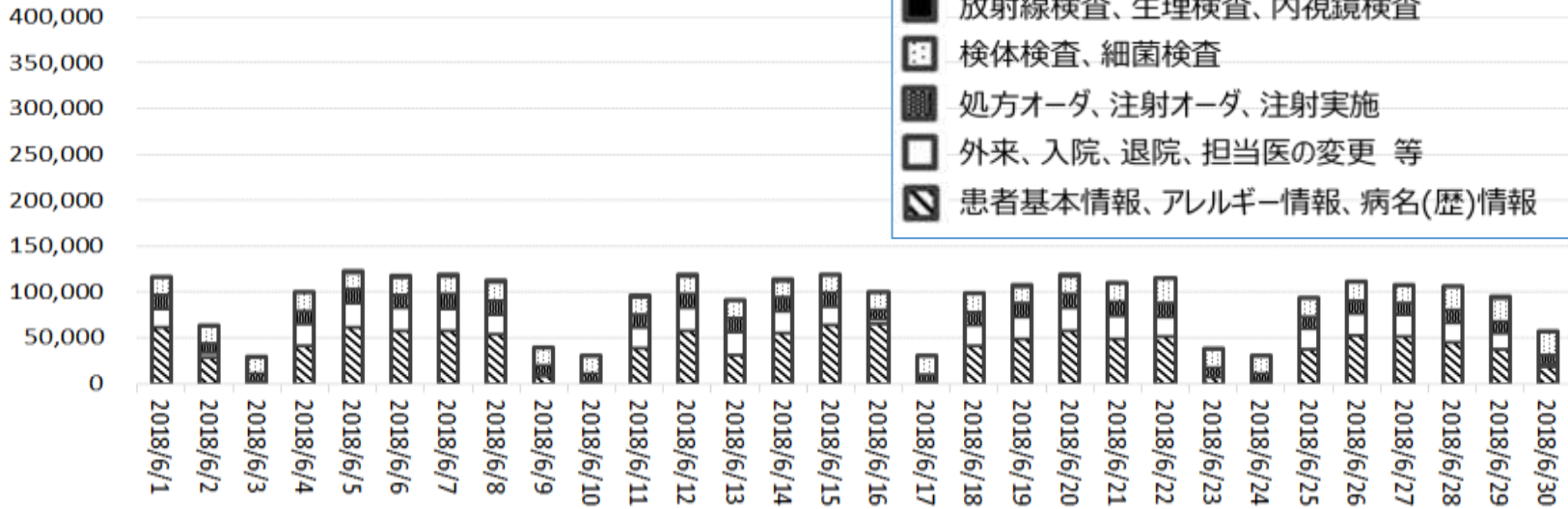
問題があった場合には運用保守業者に修正を依頼。ほぼ100% 一致

※ただし、取込み対象外データや抽出タイミングの違いによる差分は除く。  
PMDAは、個人を特定するような情報へはアクセスしていない。

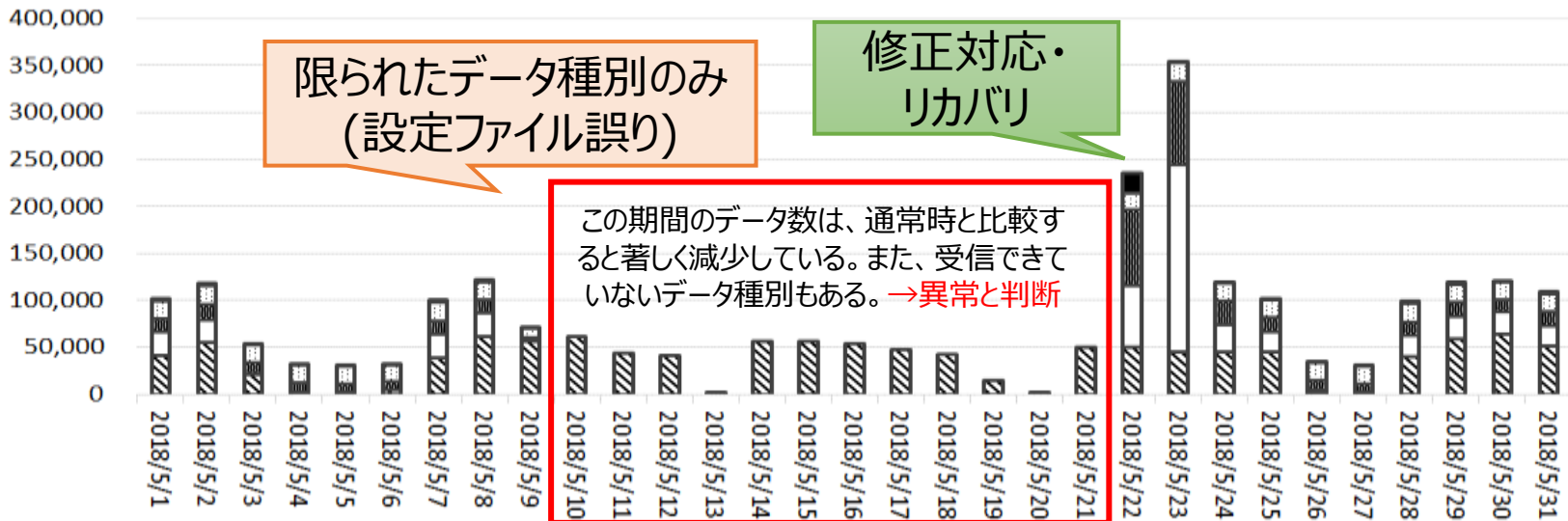


# 日常的品質管理の例 (SS-MIX 2 標準化ストレージでの受信件数)

通常時



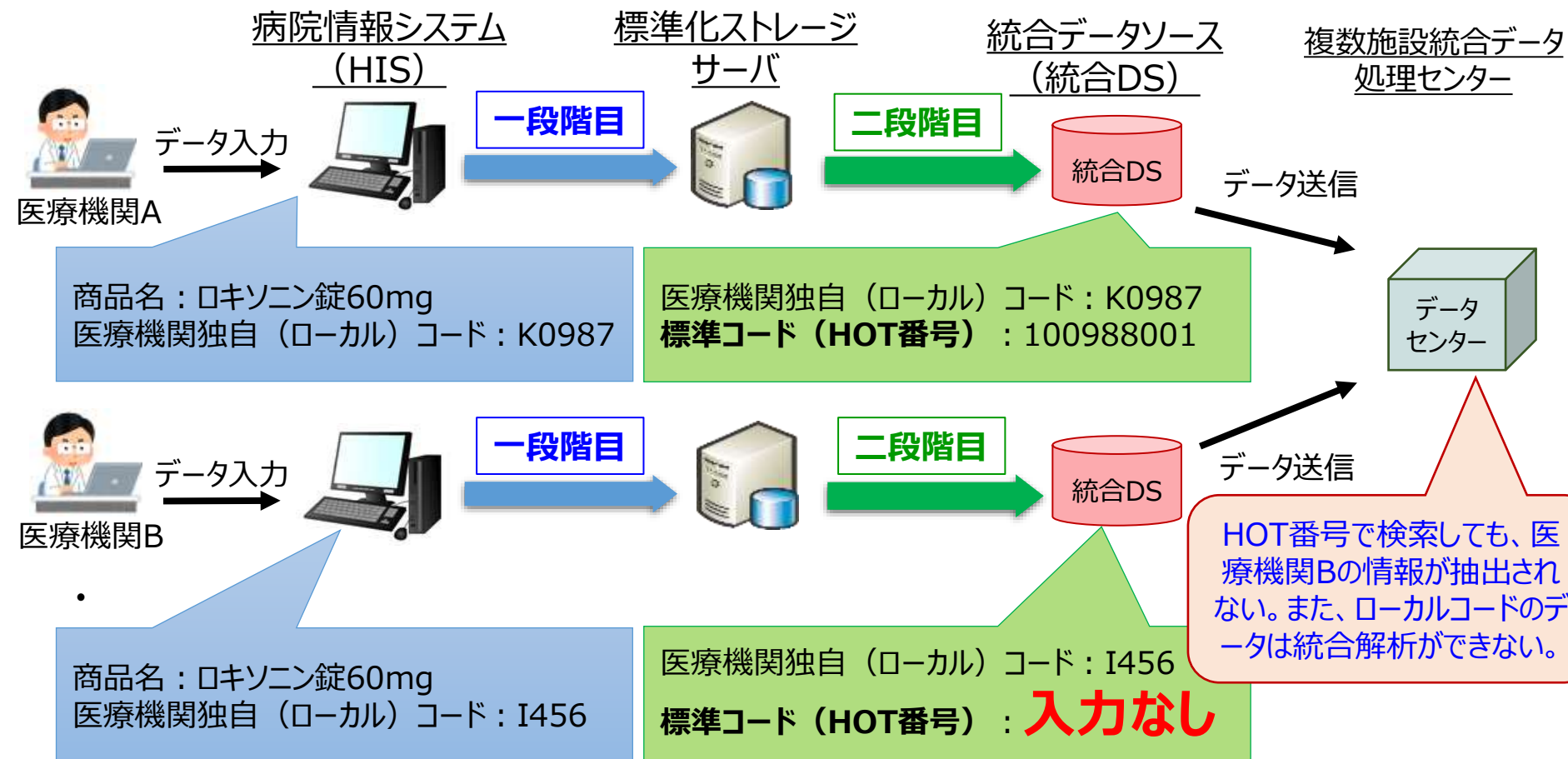
異常時



Yamaguchi, M. et al. *Pharmacoepidemiol Drug Saf* 28: 1395-1404, 2019 DOI:10.1002/pds.4879

# MID-NET<sup>®</sup>におけるデータ標準化（1）

全ての医療機関が標準コード管理できているわけではない。  
 複数病院のデータを統合して解析するためには、統一的なデータ標準化が必要。  
 MID-NET<sup>®</sup>は、統合データベースのデータ取り込み時にデータ標準化を実施。  
**※ローカルコードと標準コードの変換表の継続的な維持管理を実施。**

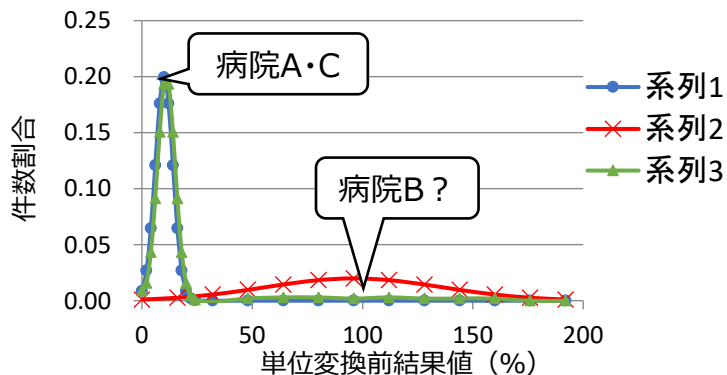


# MID-NET<sup>®</sup>におけるデータ標準化（2）

- 検体検査については、「検体検査実施頻度」及び「安全対策の観点」を考慮し、優先の高い検査を標準化対象として選定している（2023年12月末時点：**358項目**）
- 定量の**結果値換算**や**単位の統一化**も**実施**している。また、データ標準化の適切性確認のため、結果値については分布図を作成・確認している。
- 必要に応じて協力医療機関へ調査を実施し、適切な標準コードを決定している。

## 検体検査項目ごとの分布図を確認 （定量検査項目）

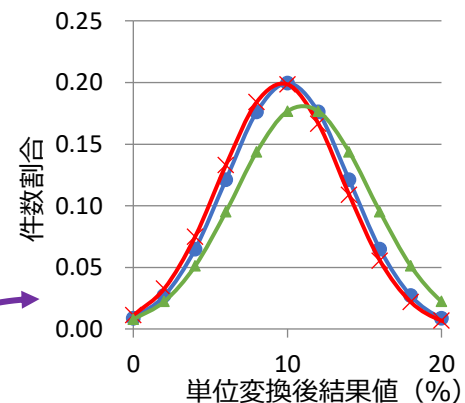
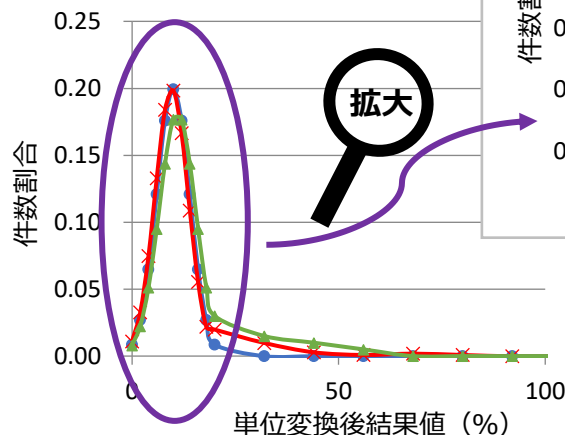
JLAC10ごとに各協力医療機関における結果値件数の合計に対する各結果値の件数割合（%）を描画



病院A・病院Cのローカル単位：%  
病院Bのローカル単位：% ? ← 検体検査件数分布を比較

## 各協力医療機関へ疑義事項を確認

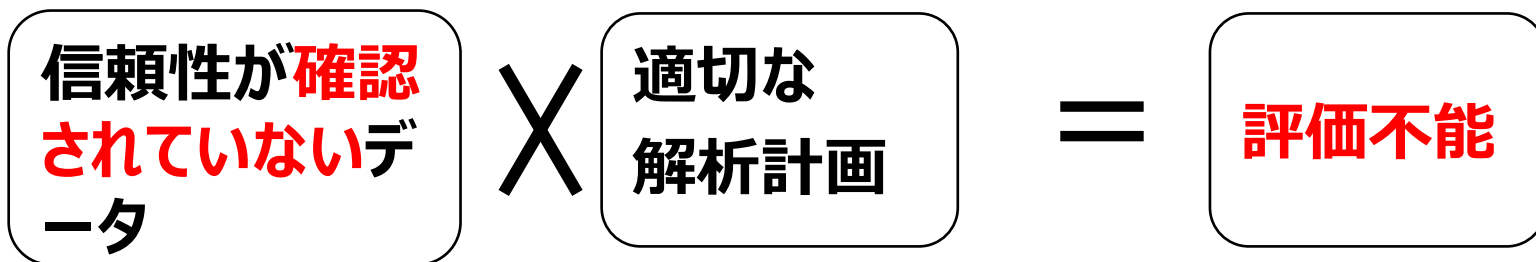
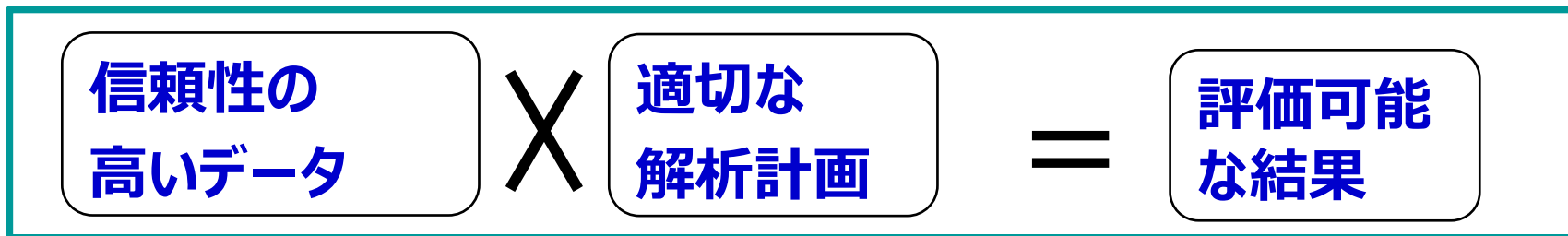
病院Bにローカル単位について  
問い合わせ  
誤：%、正：%  
⇒単位変換をして再描画



分布傾向を確認することにより、「データ標準化」の適切性を確認

# RWD活用に基づく適切な評価のために

科学的に適切で評価可能な解析結果を得るための2大要素



いくら解析計画が適切でも、データベースの信頼性が担保されていないならば、結果を適切に評価出来ない

## (1) 品質管理

- **データの信頼性が確保されていなければ、データの二次利用は困難である。**
- 電子カルテは、全国统一されておらず、同一ベンダー製であっても医療機関の事情によりカスタマイズされている。電子カルテ更新、部門システムの変更、運用方針変更等があるため、データ送信状況は常に変化する。→**医療DXによる統一に期待！**
- 現在の状況において、データベースの信頼性を確保するためには、システム稼働前に一度確認するだけでは不十分であり、継続的な確認が必要である。
- MID-NET®ではMRDA手法により**継続的な品質管理作業を実施**している。

## (2) 標準化

- **データが標準化されていなければ、データの二次利用は困難である。**
- 医薬品、検体検査等のコードは、多くの医療機関で、標準コードで管理されておらず、ローカルコードで管理されている。→**医療DXによる統一に期待！**
- MID-NET®では実態に沿った**継続的な標準化作業を実施**している。

- 今回の検討に際し、MID-NETの構築・運営時に得られた標準化、品質管理等に関する技術・知見について情報共有が可能。
- 今回の検討は厚生労働省所管DBが対象であり、MID-NETとデータ収集経路が異なる部分もありますが、MID-NETの手法（考え方）を参考に、それぞれのデータ収集方法に適した手法を検討することが重要。



# 医療等情報の二次利活用に関する技術作業班

## 医療等情報の二次利用に関する課題と解決・改善に向けて — 手法・技術・標準の観点から —

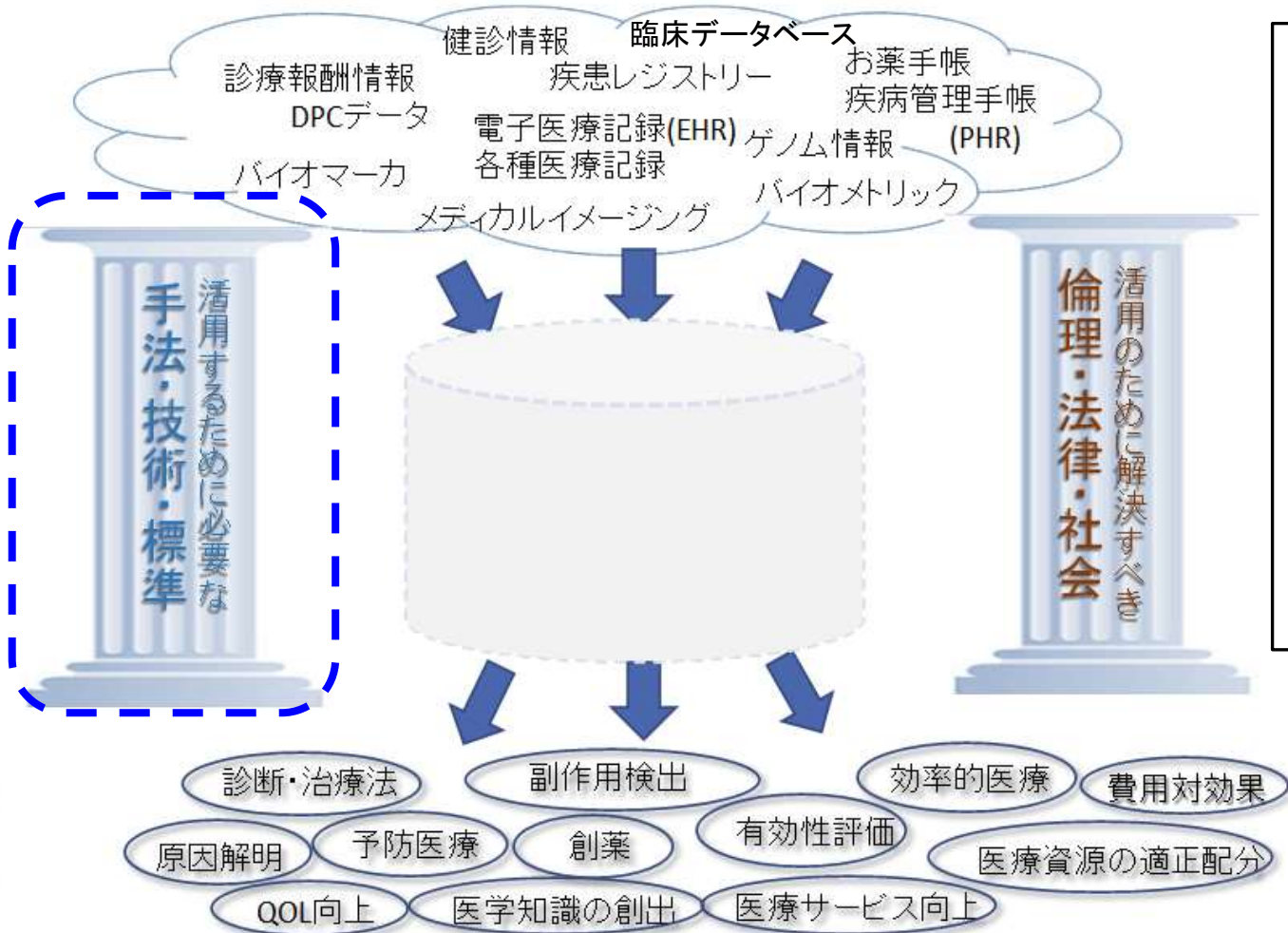
令和6年2月8日

一般社団法人医療データ活用基盤整備機構

岡田 美保子

# 日本における医療等情報の二次利用の課題

データソース



1. 医療等情報データ二次利用のために解決すべき課題には、
  - 手法・技術・標準
  - 倫理・法律・社会の両方の観点で検討する必要がある。
2. 本資料では、主に「手法・技術・標準」の観点から、医療等情報の二次利用の課題及び解決策等を提示する。



# RWD：国際的に合意された、ただ一つの定義が存在するわけではない

国際機関・海外規制当局	リアルワールドデータ（RWD）定義
ISPOR (国際医薬経済・アウトカム研究学会)	RWDとは、意思決定に利用される、伝統的なRCTs（ランダム化比較試験）により収集されたものではないデータ
ABPI（英国製薬工業協会）	RWDとは、通常の診療業務で起きていることを記述する、非介入の方法で得られるデータ
FDA（米国食品医薬品局）	RWDとは、様々なソースから日々収集される（routinely collected）患者の状態や診療の提供に関するデータ ※The Framework for FDA’s Real-World Evidence Program (FDA, Dec 2018): 医薬品の安全性と、限定的範囲での有効性（ラベリング情報の適応追加・変更、用量や用法、投与経路の変更等）の評価におけるRWEの活用が対象

RWD（リアルワールドデータ）の様々な定義等の共通特性として：

- RCTsにより収集されたものでなく、日常的な患者状態や提供される診療に関するデータ
- 非介入的な方法を用いて、観察的に得られるデータ
- EHR（Electronic Health Record：電子健康記録※）、管理的データ、医療費請求に係るデータ、セルフモニタリング機器から得られるデータや患者レジストリー等がデータソースとなる。
- RWDのデータ標準の議論では、ヘルスケアと臨床研究・薬事規制領域にわたる取り組みが必要。
- 「データの標準」と「データの品質」との関係性を明確にすることが必要である。

【※】EHR（Electronic Health Record）とは、国際標準化団体での議論（2003年頃）において、一人の患者のケアは1施設内・1臨床家に留まるものでないことから、①継続的医療（Continuity of care）を支える、診療を様々な角度から支援、②医療者、患者、家族を支援、③医療の質向上、ポピュレーションヘルスに資する、他施設との医療データ交換に対応し得る（相互運用性を有する）システムである。→ この議論から20年が経過しているが本質は現在も変わらない。

※医療DXの時代では一次利用と二次利用の境界は曖昧になっている。

## EHRの目的

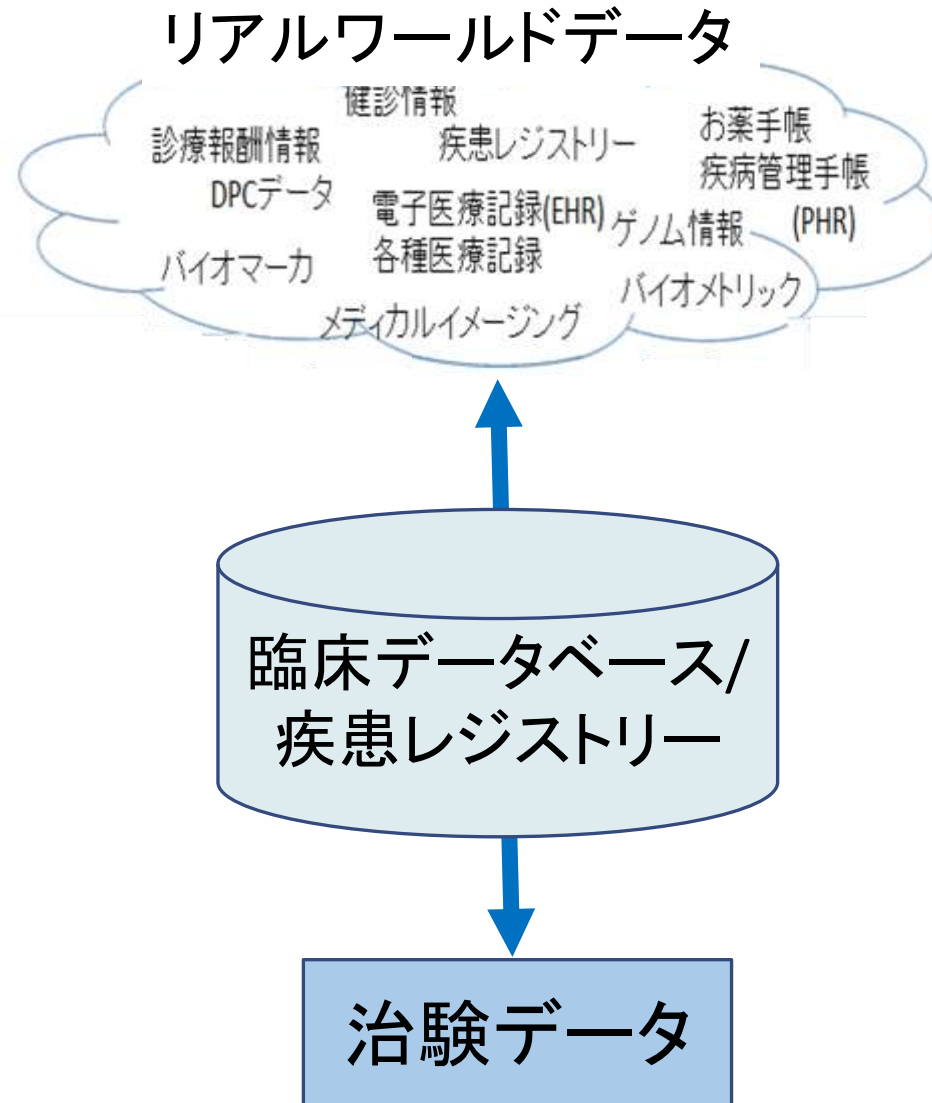
ケアの継続性  
ケアのコーディネーション

医療の質向上  
ポピュレーションヘルス

# RWD（リアルワールドデータ）の種類

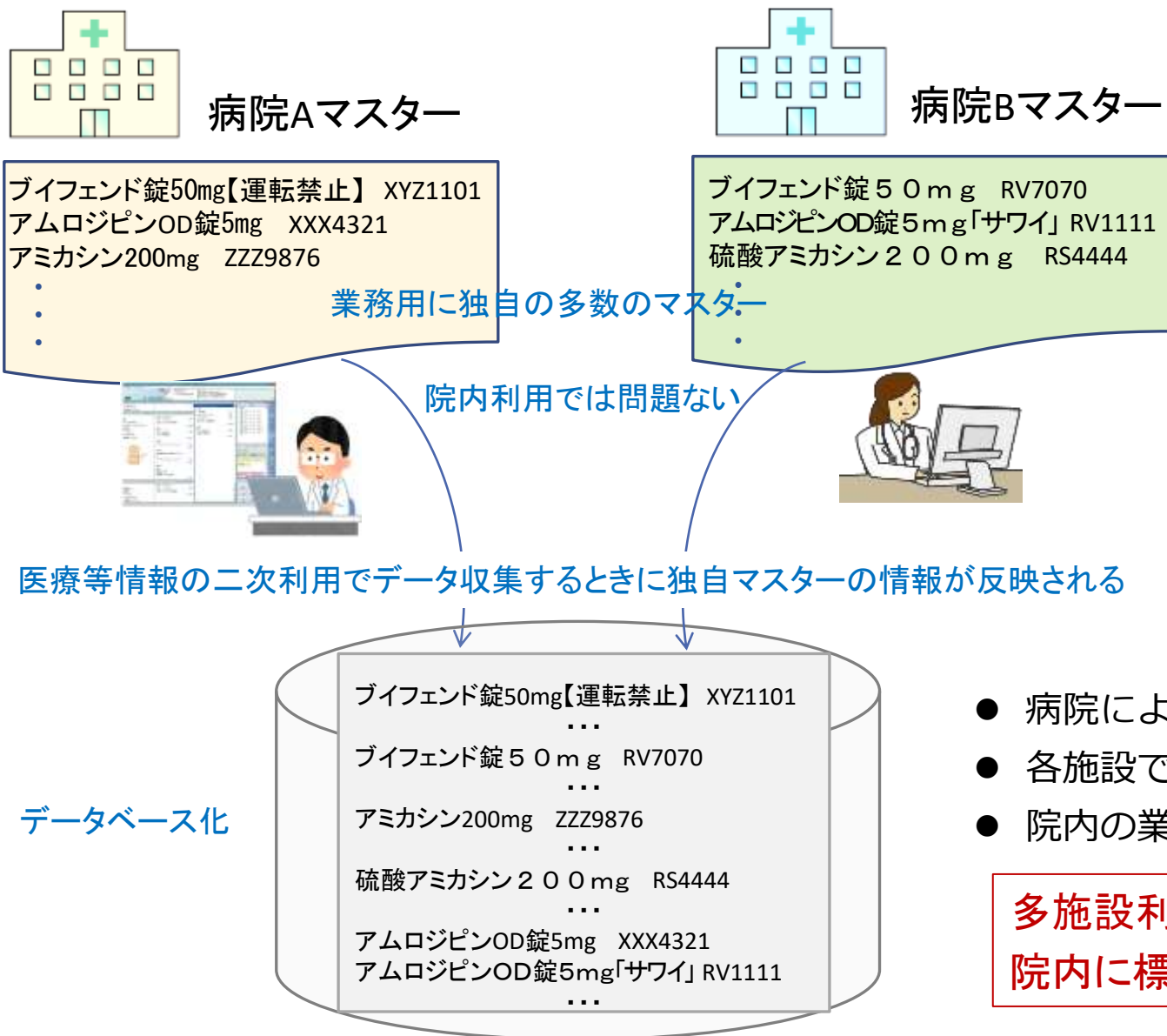
## やや広い範囲のRWD（リアルワールドデータ）※

	研究目的	非研究目的				
	試験データ	レジストリー/DB	日々収集されるRWD 議論の中心はこの範囲			調査データ
一次目的	治験、臨床試験	目的は多様	医療記録 (電子カルテ)	DPC、レセプト等	セルフモニタリング、PHR、ウェアラブルデバイス等	国民調査 病院調査 患者調査等
データ収集	プロトコルに則った収集	患者レジストリー、臨床DB、ゲノムDB、etc.	日々の記録	所定の様式	ソースにより多様	調査様式
二次目的	従来適用されないことが多い	観察研究利用				
データ品質基準	国際的な基準あり	様々な議論あり。規制領域では限定された範囲でのガイドラインあり。国際的に合意されたRWD全般に適用可能な基準はない				

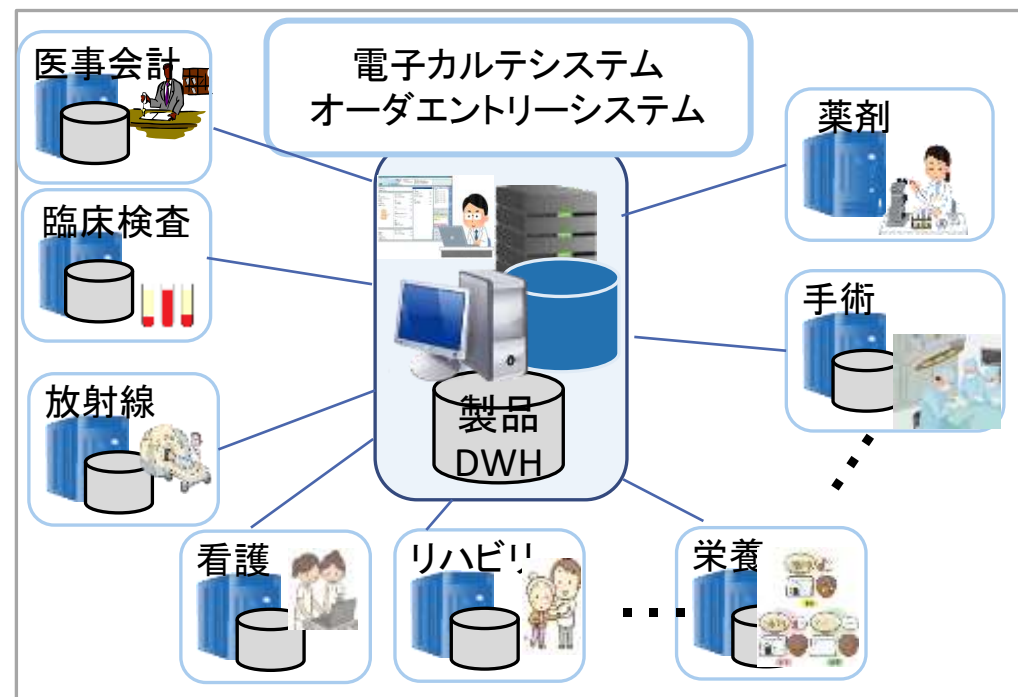


※いずれかの一つの公式定義によるものではない。

# 病院情報システムは院内業務を支援する



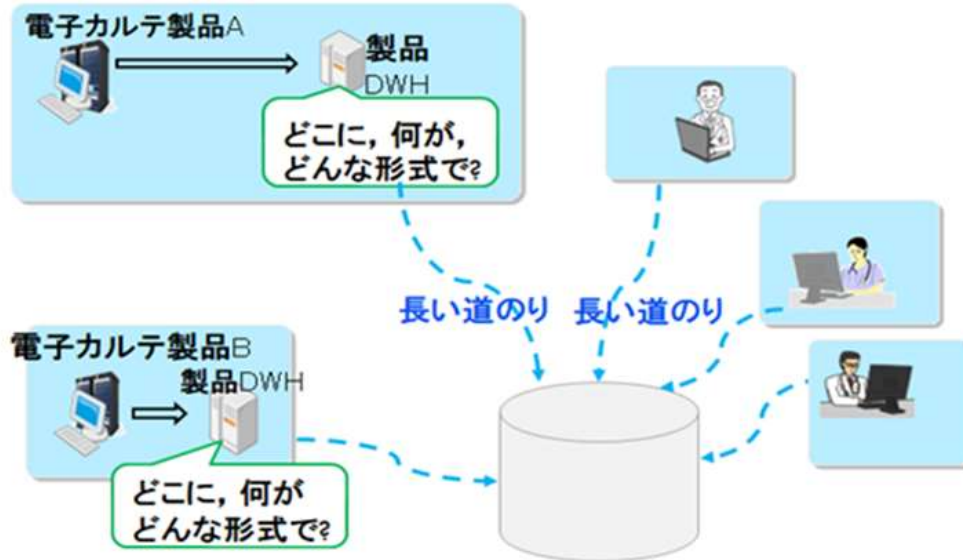
## 病院情報システム



- 病院により部門システムの種類、ベンダー、導入時期など異なる。
- 各施設で独自カスタマイズ、独自マスター作成・運用ルールがある。
- 院内の業務用の目的に独自マスターが作成されている。

多施設利用には標準化されたマスター整備が必要である。  
院内に標準を取り込む現実的な解決策を模索している。

## 多施設からの診療データの収集



電子カルテ製品ごとに、何がどこに、どのようなデータ形式に対応しているのか異なる。

- データ入力が手入力に頼らざるを得ない
- 臨床家の入力負担が増して手入力に限界がある。
- 都度データ収集、整備に膨大な時間・費用・労力

診療情報の利活用が進むとともにデータ入力の手間やシステムの整備に様々な問題が顕在化している。

## 様々な収集方法・解析方法

### 【データベース/レジストリーの構築】

- ✓ ウェブからの手入力
- ✓ 電子カルテのテンプレートからの手入力
- ✓ SS-MIX2を用いた自動収集
- ✓ SS-MIX2からの自動収集と電子カルテテンプレートの手入力のハイブリット
- ✓ DWH (Data Ware House: データウェアハウス) による自動収集

### 【分散型のデータ利活用】

- ✓ オンデマンドによるデータ収集・解析
- ✓ 分散型のデータ解析

# 診療データ収集方法の事例

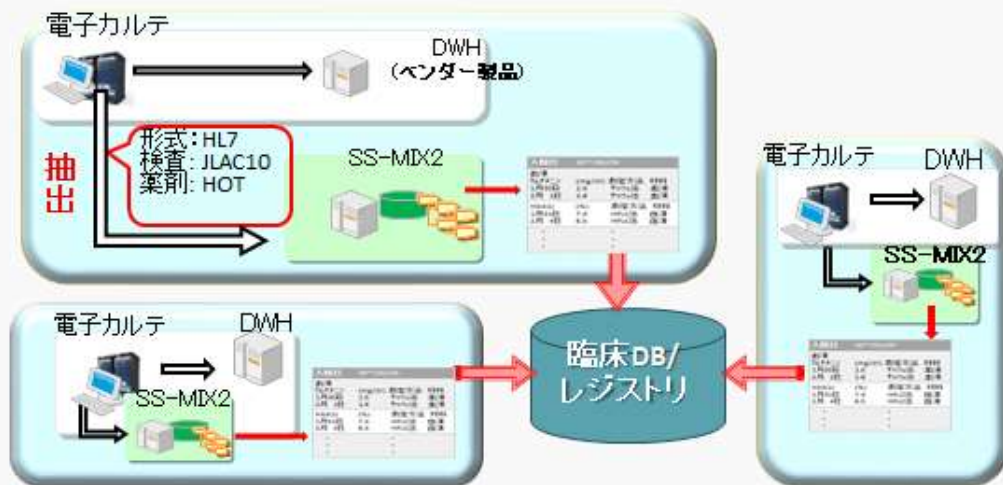
【例1】Webで手入力



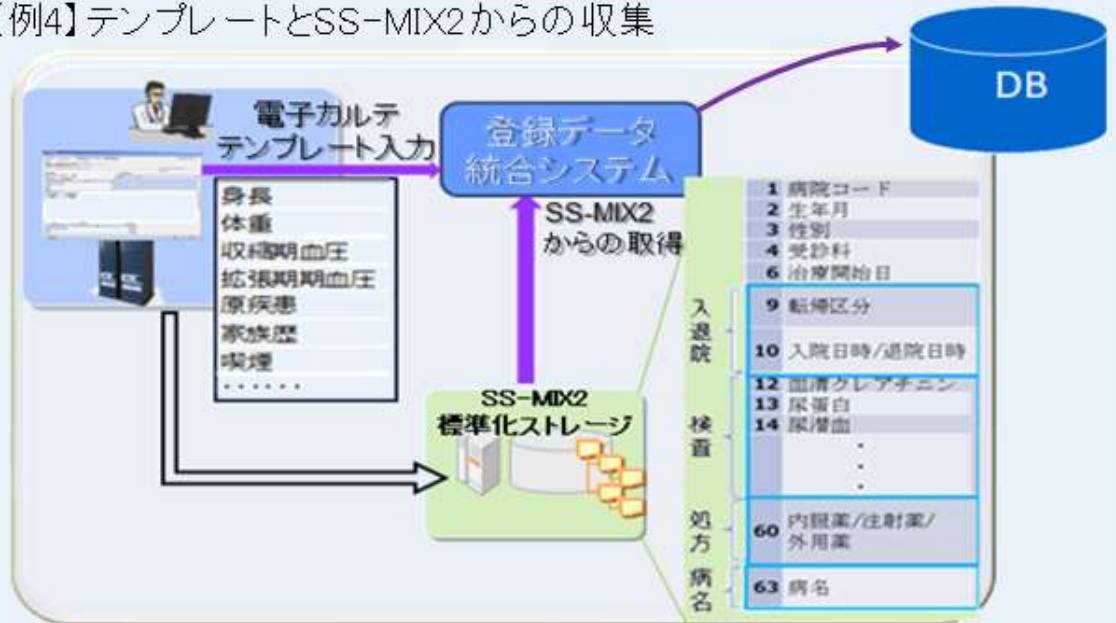
【例3】電子カルテテンプレートからの入力



【例2】SS-MIX2からデータ収集



【例4】テンプレートとSS-MIX2からの収集



# 国内における医療データベースの類型

	臨床レジストリー/データベース	レギュラトリーレジストリー/データベース	公的レジストリー/データベース
目的	疾患、治療、患者状態、アウトカム、等を扱う臨床研究	医薬品・医療機器の市販後安全性調査等	保険制度の評価、国民、地域住民の健康状態把握、医療提供体制の評価、等
具体的活用目的	医療の質の評価 有効性（臨床的効果等）評価 効率性（費用対効果等）評価 アウトカム評価 治療の満足度 疾病の自然経過、等 患者支援、患者の研究参画支援	市販後サーベイランス 安全性モニタリング ハイリスクデバイスの監視 シグナル検知 有効性評価 安全性評価、等	感染症等 疾患サーベイランス 公的保険診療の政策 特定疾患・指定難病等 出生、死亡 出産、流産
構築・運営主体	学術団体、研究機関、研究者グループ、患者支援団体を含む組織 等	企業、国、行政機関、研究機関 等	国、自治体、研究機関 等
国内の事例	NCD: 一般社団法人National Clinical Database (日本外科学会等の外科系諸学会) NCDA: 国立病院機構診療情報集積基盤 JROAD: 日本循環器学会循環器疾患診療実態調査 J-DREAMS: 国立国際医療研究センター・日本糖尿病学会 J-CKD-DB: 日本腎臓学会包括的腎臓病データベース J-DOME: 日本医師会かかりつけ医診療データベース研究事業	MID-NET	NDB (匿名医療保険等関連情報データベース) 介護保険総合データベース (介護DB) DPCデータベース 全国がん登録 データベース 指定難病患者データベース 小児慢性特定疾病児童等データベース

国内における主な公的データベース	顕名/匿名
全国がん登録 データベース	顕名
指定難病患者データベース	顕名
小児慢性特定疾病児童等データベース	顕名
NDB（匿名医療保険等関連情報データベース）	匿名
介護保険総合データベース（介護DB）	匿名
DPCデータベース	匿名

## 公的データベースの二次利用における課題

### 既存の公的データベース活用の困難性

- NDB等の国が保有する大規模な公的データベースの利用は限定的である。
- 死亡情報の利用にも課題がある。

### 公的データベースと民間データベース等の同士のデータベース同士のデータ連結が困難。

- 標準規格及び標準コードが統一化されておらず、識別子によるデータ連結が困難。

**法律的な観点からの議論も必要**



**公的データベースと次世代医療基盤法の認定事業者が保有するデータベースの連携**

# 医薬品コード（薬剤データ）の利活用に係る課題

- 薬事や流通では一定の規制や産業の枠組みのもとで医薬品コードが利用されている。
- 処方では医薬品コードの標準の位置づけが曖昧でリアルワールドデータを用いた研究に課題がある。

## ■ 病院で利用される医薬品コード

- 流通用コード（GS1）、レセプト用コード（レセプト電算コード）、電子カルテ用コード（YJコード、ローカルコード）と目的別の医薬品コードがある。
- レセプト用コード（レセプト電算コード）は制度、流通用コードは業界デファクトとして運用されている。
- 電子カルテの処方マスタには病院固有のローカルコード、YJコードが使われている。

## ■ 厚生労働省標準規格として「医薬品HOTコードマスター」が指定されている。

- 病院業務に「HOTコード」は利用されておらず、導入が進まない。
- 厚労省事業、厚労科研では厚生労働省標準規格「HOTコード」の使用が求められている。
- 臨床DB構築では事業・施設ごとに「HOTコード」に変換し、手間・労力の問題及びエラーのリスクもある。



手動変換を必要としない標準コードが流通する仕組みが必須である。



# 検体検査データの活用上の課題

- 検査データの単位は施設によバラバラ、表記にもかなりのばらつきがある。  
(例) WBCの単位 x1000/ $\mu$ L、10S3/ $\mu$ L、 $\times$ 10S2/ $\mu$ l、 $\times$ 千/ $\mu$ L など
- 施設ごとの検査値の分布
  - 長期の蓄積データでは途中の基準値の変更が把握できない。
  - 基準値が揃っていても検査値の分布が施設で異なる場合もある。
  - 多施設から収集される検体検査データの検査項目の表現にばらつきがある。
  - 蓄積データからは、基準範囲（正常範囲）を把握できない場合がある。
- 定性値が標準化されておらず、検査値の表記揺れが大きい。
  - 例えば尿蛋白（定性）は、病院により「 $\pm$ , 1+ $\sim$  3+」や「+1 $\sim$ +5」など
  - データとしては、+-、 $\pm$ 、(+1)、1+、(1+) 30、全角数字・記号等の様々なばらつきがある。
- 検体検査項目の識別上の課題
  - 標準のJLAC10コードは、医療施設にほとんど普及していない。
  - 割当が不完全であるので、DB上でJLAC10コードとローカルコードが混在する。
  - JLAC10コードで完全に確定できない場合があるので、JLAC11コードへの移行を進めている。
  - 検査項目の名称は施設により表記が異なる。

# 電子カルテ情報の活用に係る課題

## 課題

### ① 電子カルテ情報で取得・把握が困難な例

- 医薬品（薬剤）の用法・用量の標準規格が普及していないため、医薬品（薬剤）の実際の投与量の取得・把握が困難である。
- 医薬品（薬剤）投与のリーズニング（投与開始、投与継続、投与中止等の理由）の取得・把握が困難である。
- 電子カルテに記載されているが、正確な取得・把握が難しい項目  
（例）体重（更新されていないことが多い）、血圧値（入力方法が定まっていない）
- 患者の状態に関する情報、詳細な病態に関する情報（ナラティブに記載されている）
- 予後情報・死亡情報を取得することが難しい。

### ② 電子カルテ情報が標準化されておらず、直接的に取得・把握することが困難な例

- 疫学研究等で使用されるが、患者プロフィールで標準化されていない項目  
（例）アレルギー、予防接種、手術歴、家族歴、社会歴、喫煙等
- プロブレム、プロブレムリスト
- 用語の意味が標準化されていない項目  
（例）イベント、転帰情報
- DPCにはあるが、電子カルテ情報では標準的に管理されていない項目  
（例）ADL、手術情報
- その他、システム運用上の違いによりデータ上で問題が生じかねない情報

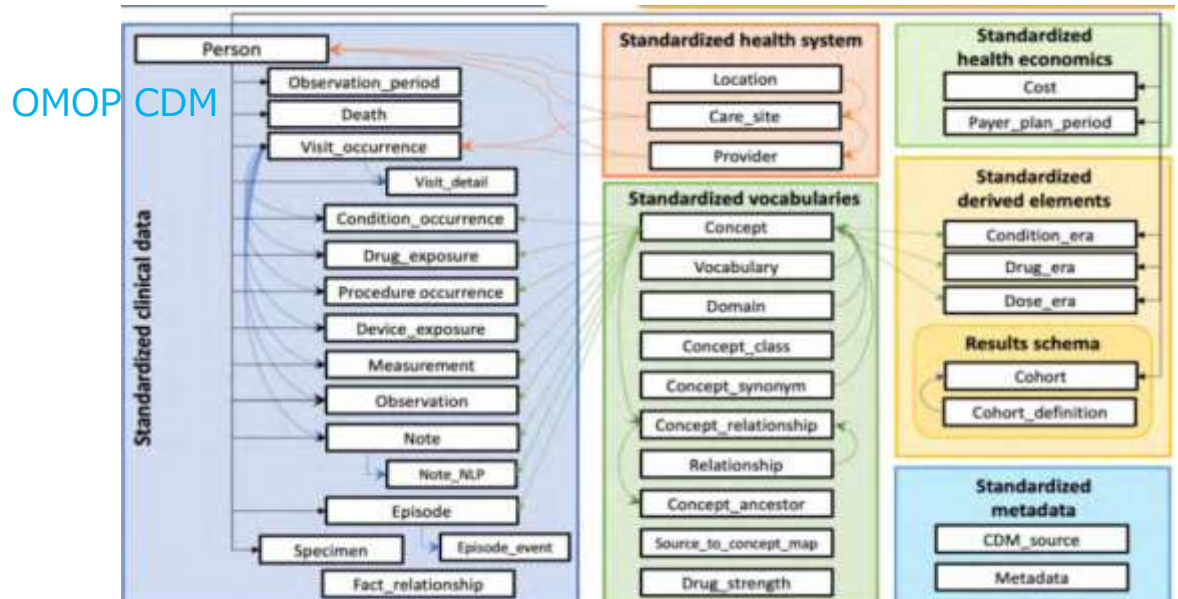
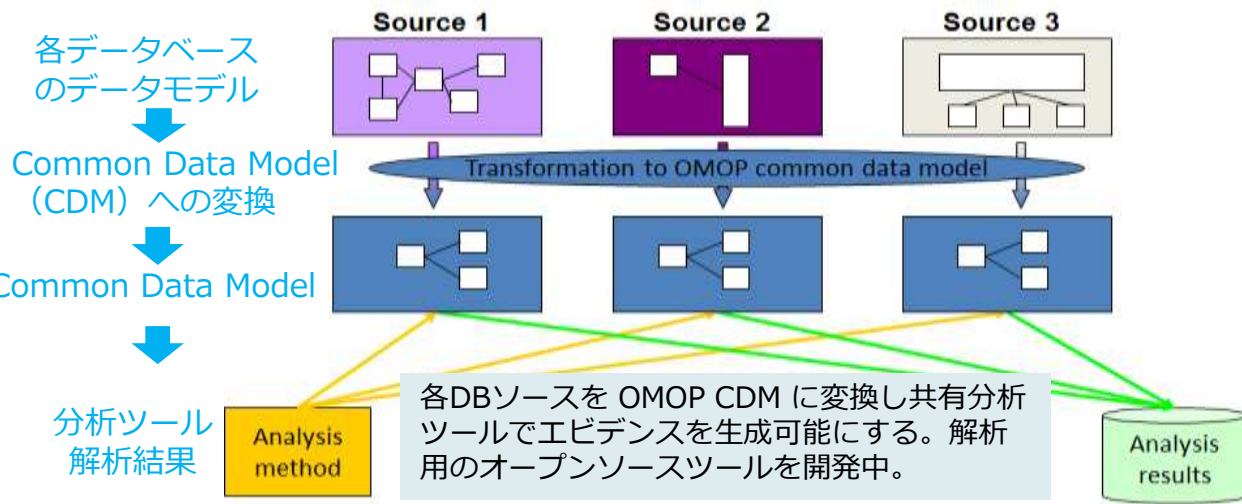
# RWDの二次利用 – データ連結に係る全体的課題



全般的課題	課題	解決策/改善策の例
DBの横展開・データの統合解析	<ul style="list-style-type: none"> <li>個々の事業のデータベースには共通性がなく、標準的データを揃えられない。</li> </ul>	<ul style="list-style-type: none"> <li>異なるアプリケーションやデータソースでデータを交換できる標準化データモデルである「コモンデータモデル (Common Data Model : CDM (※))」を検討する。</li> <li>※CDM : データベースがどのような情報から構成されるのかをデータモデルとして記述する。複数のDBに対して一つのコモンデータモデル (CDM) を定義し、個々のDB側のデータモデルをCDMに変換し、データ解析等を可能とする。</li> </ul>
標準化されていないRWD	<ul style="list-style-type: none"> <li>各事業でローカルコードから標準コードへの変換を実施しているが、作業の重複及びデータの信頼性等の課題がある。</li> </ul>	<ul style="list-style-type: none"> <li>各種のデータ標準及び標準コードのマスターの整備及び管理等を一元的に行う組織・団体等を設置する。</li> </ul>
	<ul style="list-style-type: none"> <li>標準コードの付与に関して、誰がどのように、どのタイミングでやっていくのか、コード付与等の実装ガイダンスが一般化されていない。</li> </ul>	<ul style="list-style-type: none"> <li>RWDを用いたDB構築事業等に対する実装ガイダンスを策定する。</li> </ul>
公的DBと民間DBのデータ連結	<ul style="list-style-type: none"> <li>民間DBとNDB等の公的DBとデータを連結できれば、研究の幅、量、質とも各段に向上することが期待される。</li> <li>民間DBと公的DBを連結させる「キー情報」が必要である。</li> <li>長期のアウトカムの評価や治療成績などの研究が行えない。</li> </ul>	<ul style="list-style-type: none"> <li>RWD側で「被保険者番号」が入っていれば、支払基金等が運営している「被保険者番号」の履歴データベースにアクセスして同一性を確認できる。</li> </ul>
妥当な連結	<ul style="list-style-type: none"> <li>国レベルの大規模データベースと電子カルテシステムを用いたレジストリー・DBの連携では、ポピュレーションの特定が必要となる。</li> </ul>	<ul style="list-style-type: none"> <li>対象とするDBのメタ情報を整備してはどうか。</li> </ul>
国際共同研究	<ul style="list-style-type: none"> <li>国際共同研究では国際標準が必要となる。</li> </ul>	<ul style="list-style-type: none"> <li>CDMを採用する方法を検討してはどうか。</li> <li>OHDSI OMOPも参考となる。</li> </ul>
収集すべきデータ項目	<ul style="list-style-type: none"> <li>手術や治療行為に関する情報などは、レセプト情報やDPCから取得できるが、多施設共同研究の場合、施設により活用しているシステム等が異なるので、その利用に時間を要する。</li> </ul>	<ul style="list-style-type: none"> <li>国が保有・管理する公的DB等からデータを取得できるようにする。</li> </ul>
	<ul style="list-style-type: none"> <li>臨床研究・疫学研究に重要な「予後情報」は取得できない。</li> <li>院内にあるがん登録の死亡情報は研究利用できない。</li> </ul>	<ul style="list-style-type: none"> <li>縦断的データについては、公的データベースの議論を参照する。</li> </ul>

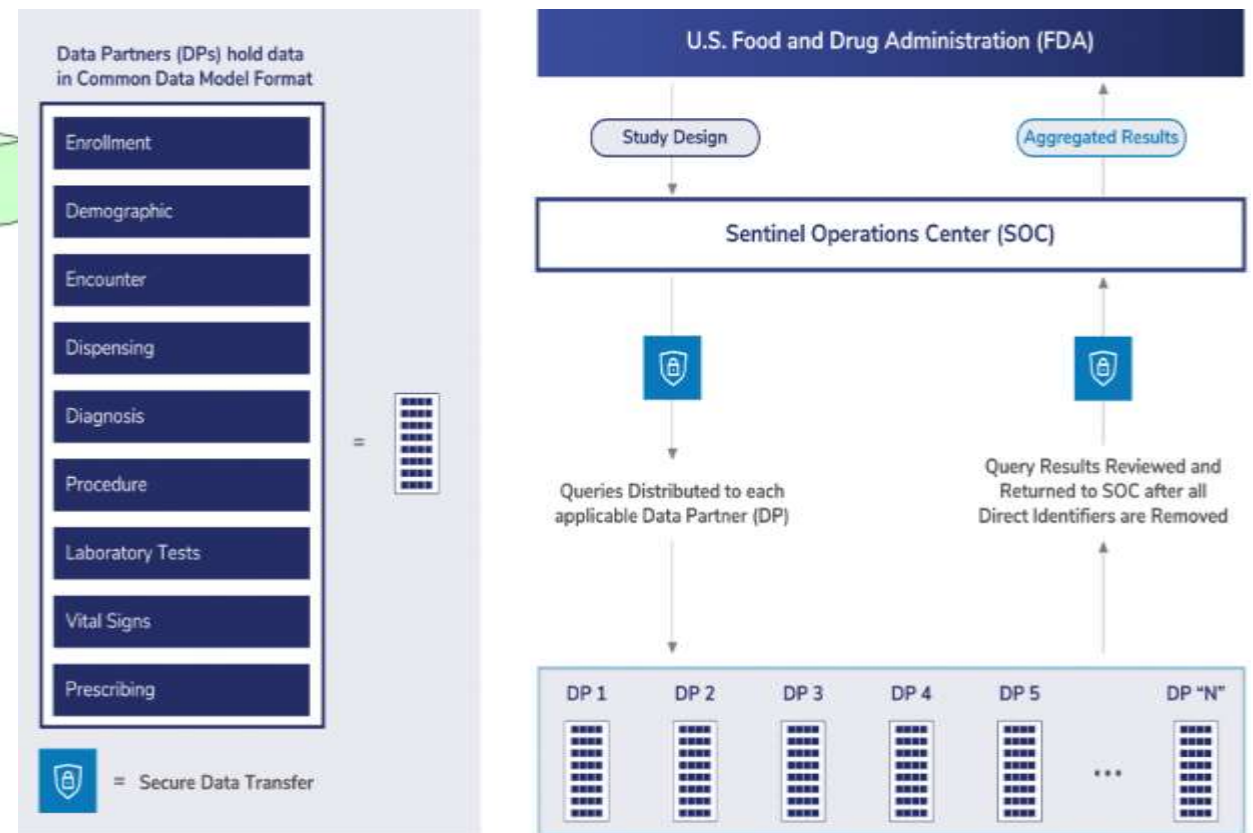
# コモン・データモデル (Common Data Model: CDM) の例

【例1】 観察研究の国際共同研究OHDSIの「OMOP CDM」  
 (The Observational Medical Outcomes Partnership Common Data Model)  
<https://www.ohdsi.org/data-standardization/>



【例2】 SENTINEL Common Data Model (米国FDA)  
<https://www.sentinelinitiative.org/methods-data-tools/sentinel-common-data-model>

- 【SENTINEL Common Data Model (米国FDA)】
- 米国FDA承認の医薬品及び医療製品の安全性モニタリングするための長期プログラム
  - 医療費請求情報とEHRを有するデータパートナー (DP: 医療組織) で構成。
  - データは定期的に収集、各DPは各自のデータを保持、誰がアクセスできるかを制御。
  - プログラムで分析を実行し、匿名化された結果をオペレーションセンターに送信。
  - 患者のプライバシーとデータセキュリティを確保。
  - 「CDM・分散アプローチ」は、稀な有害事象や少数集団に適用される薬剤の研究ができるよう大規模データセットを作成する必要がある。標準化されたデータ構造としてCDMを定義。DPはローカルでSENTINEL CDMに変換し、識別可能な患者情報を削除・マスクする。



# RWDの二次利用 — 医薬品（薬剤）・検体検査の標準化に係る課題

医薬品（薬剤）	課題	解決策/改善策の例
厚生労働省標準規格	<ul style="list-style-type: none"> <li>厚生労働省標準規格では「HOTコードマスター」（HOT基準番号を他のコードと対応付けたもの）を指定しており、厚労省事業、厚労科研等では厚生労働省標準規格に準じる必要がある。</li> <li>「HOTコード」は院内業務用ではなく、院内で活用されていない。</li> <li>厚生労働省DB事業では「HOTコード」に変換して保存する必要がある。</li> <li>医療従事者の負担、エラー、継続的な標準コードの維持管理に課題がある。</li> </ul>	<ul style="list-style-type: none"> <li>医療施設内で確実に導入される標準コードを推奨するため、厚生労働省事業で「YJコード」を利用できるようにしてはどうか。</li> <li>薬機法改正で添付文書と「GTINコード」が紐付けられたので、GTINコードと他の標準コード（YJコード、HOTコード）を紐付けたマスターの整備を進めてはどうか。</li> </ul>
用法・用量・単位	<ul style="list-style-type: none"> <li>RWDでは一般に医薬品（薬剤）の投与量を統一的に取得できない。</li> <li>電子処方箋の医薬品の用法・用量等について議論がなされている。</li> </ul>	<ul style="list-style-type: none"> <li>医薬品の用法・用量・単位の標準の確立及び普及を図ってはどうか。</li> </ul>
検体検査	課題	解決策/改善策の例
JLAC10 JLAC11	<ul style="list-style-type: none"> <li>検体検査結果項目コードは施設内でローカルコードが使用されている。</li> <li>検体検査項目コードの厚生労働省標準規格はJLAC10コードが採用されているが、医療機関の施設内ではほとんど導入されていない。</li> <li>既存のRWD/臨床DBではJLAC10コードの割当が多い。</li> <li>医療機関の臨床検査部ではJLAC10コードの正確な割当が相当困難である。</li> <li>JLAC10コードからJLAC11コードに移行しつつあるが、既存DBの検査項目にJLAC11コードを全て割当てるとは相当困難である。</li> </ul>	<ul style="list-style-type: none"> <li>JLAC11コードはJLAC10コードの課題を解決したコードであり、新規の体外診断用医薬品の場合はJLAC11コードを付番する方針してはどうか。</li> <li>対応策として、「マスター対応表」の構築・公開を行ってはどうか。 <ul style="list-style-type: none"> <li>- 複数DB間で共通する検査項目にJLAC10とJLAC11の両方を割当てる。</li> <li>- 既存DBのJLAC10コードに全てJLAC11を採番する。 <ul style="list-style-type: none"> <li>※付番：新規の対外検査診断医薬品にJLACコードを当てること。</li> <li>※採番：既存の院内検査項目等にJLACコードを割り当てること。</li> </ul> </li> </ul> </li> </ul>
JLAC11の現状	<ul style="list-style-type: none"> <li>JLAC11コードは現時点で280項目のみが公開されている。</li> </ul>	<ul style="list-style-type: none"> <li>JLAC11で付番されていて、公開されていないものは早急に公開する。</li> <li>JLAC11が付番されていない項目は、付番委員会でただちに付番する。</li> </ul>
JLAC10/11コードの付番・採番・認証に係る組織体制	<ul style="list-style-type: none"> <li>JLAC10/11コードに係る全体的な判断、質問・要望する先の機関・組織が確立していない。</li> <li>MID-NETではJAC10コードが空欄で差分抽出されたものをガバナンスセンタで割当・返却しているが、この継続は費用対効果から困難である。</li> </ul>	<ul style="list-style-type: none"> <li>日本全体でデータが使えるためには管理・付番、採番し、日本全体に普及させ、オーサライズできる非営利の組織及び仕組みを確立する。</li> </ul>
院内システム対応	<ul style="list-style-type: none"> <li>DB登録時にJLAC10/11コードを割り当てる方法では他の用途に利用できず、維持管理が困難である。</li> </ul>	<ul style="list-style-type: none"> <li>院内システムの上流工程で、JLAC10/11コードが入る仕組みを作り、一次利用、二次利用ともに標準化を図る。</li> </ul>
過去データ	<ul style="list-style-type: none"> <li>長期のデータに関しては、方法、機器、試薬等が変更される場合に値が影響するが、現状のDB上にこれらの情報は無い。</li> <li>過去に遡って院内システムから取得したデータの信頼性の問題もある。</li> </ul>	<ul style="list-style-type: none"> <li>検査の仕様等を変更する場合に、医療機関等のマスターで登録するとともに、この情報をDBに反映させる必要がある。</li> <li>具体的な目的ごとに対応するのが現実的である。</li> </ul>

# RWDの二次利用 ー医薬品（薬剤）コード

医薬品コード	概要	例
薬価基準記載医薬品コード	<ul style="list-style-type: none"> <li>● 薬価基準に記載された医療用医薬品に付された英数12桁の分類コードであり、「厚生労働省コード」とも呼ばれる。</li> <li>● 医療用医薬品の記載方法には、「銘柄別記載」と「統一名称記載」の2種類があり、記載方法によってコードの付与のされ方が異なる。</li> <li>● 官報告示された薬価に対して付与されるため、異なる医薬品に同一のコードが付与されて商品が識別できない場合がある。</li> </ul>	<ul style="list-style-type: none"> <li>● □キシロニ錠60mg（第一三共）：1149_019_F_1_56_0</li> <li>● □キシプロフェンNa錠60mg「テバ」（武田テバファーマ）1149_019_F_1_65_0</li> </ul> <p>①薬効分類番号、②投与経路、③剤形、④同一分類内での規格単位の連番 ⑤同一規格に対して付与された番号、⑥チェックディジット</p> <p>※□キシロニ錠60mgと後発品は成分、剤形、規格が同じなので最初9桁が一致。</p>
レセプト電算処理システム用コード（支払基金コード）	<ul style="list-style-type: none"> <li>● 医療機関が審査支払基金に電子レセプトを提出する際に使用する9桁の分類コード</li> <li>● 医療費を請求する際に使用されるため、薬価基準記載品のみが対象となる。</li> </ul>	<ul style="list-style-type: none"> <li>● □キシロニ錠60mg：6_20098801</li> </ul> <p>① 医薬品は「6」で固定 ② 医薬品毎に一意的な番号</p>
個別医薬品コード（YJコード）	<ul style="list-style-type: none"> <li>● 薬価記載された医療用医薬品ど同一の個々の医薬品に対して付与した英数12桁で構成された分類コードである。</li> <li>● 「銘柄別記載医薬品」の場合は、「薬価基準記載医薬品コード」と「個別医薬品コード（YJコード）」とは同じコードになる。</li> <li>● 「統一名称医薬品」の場合は、YJコードは個々の医薬品に対して個別のコードが付与されるため、薬価基準記載医薬品コードとYJコードは異なる。</li> <li>● YJコードは、保険記載形態によらず品目の管理ができ、医療システム、医療用医薬品添付文書の電子化（e-添付文書）などで汎用されているコードである。</li> </ul>	<ul style="list-style-type: none"> <li>● □キシロニ錠60mg（第一三共）：1149019F1560</li> <li>● □キシプロフェンNa錠60mg「テバ」（武田テバファーマ）：1149019F1650</li> </ul>
GTINコード（Global Trade item Number）	<ul style="list-style-type: none"> <li>● GS1が標準化した国際標準された商品コードであり、医薬品もコード化されている。</li> <li>● JANコード標準型はGTIN-13、包装用商品コードはGTIN-14に相当する。</li> <li>● 医療用医薬品の調剤包装単位はGTIN-13、販売包装単位及び元梱包包装単位はGTIN-14が使用され、バーコードとして使用されている。</li> <li>● 医療用医薬品では、GTINコード（調剤包装単位、販売包装単位）に加えて、有効期限やロット番号も記載されている。</li> </ul>	<p>（例）□キシロニ錠60mg：1_49_87081_10540_0</p> <p>① 包装単位識別（0：調剤包装単位 1：販売包装単位） ② 国コード：日本（49） ③ GS1事業者コード（87081） ④ 商品アイテムコード（10540） ⑤ チェックディジット（0）</p>
HOTコード	<ul style="list-style-type: none"> <li>● 医療用医薬品が対象に付与された各医薬品コードを横断的に対応付ける13桁の分類コード。</li> <li>● 薬価基準記載医薬品コード、個別医薬品コード（YJコード）、レセプト電算処理システム用コード、GTINコードを相互に対応づける。</li> <li>● 厚生労働省標準規格であり、MEDIS-DCが管理している。</li> </ul>	<p>（例）□キシロニ錠60mg（PTP 10T×10）：1009880_01_02_04</p> <p>①処方用（1009880）②会社識別用（01）③調剤用（02）④物流用（04） ①～④を組み合わせると各医薬品コードとの対応を表現する。 ①：薬価基準記載医薬品コード ①②：個別医薬品（YJコード）or レセプト電算処理システム用コード ①②③：GTINコード（調剤包装単位） ①②③④：GTINコード（販売包装単位）</p>

# RWDの二次利用 — 傷病名コードの標準化に係る課題



傷病名	課題	解決策/改善策の例
厚生労働省標準	<ul style="list-style-type: none"> <li>厚生労働省標準として、医療情報システム開発センター（MEDIS）の「標準病名マスター」の病名及びICD10コード（国際疾病分類）が国内で普及している。</li> <li>ICD-11コードは、分類コードだけでなく、URI（Uniform Resource Identifier）を用いてより詳細な情報管理が可能であり、ICDコード、ファウンデーションID、用語IDを使用できる。</li> </ul>	<ul style="list-style-type: none"> <li>厚生労働省標準として、医療等情報の一次利用の観点も踏まえつつ、ICD10/11コードの使用を推奨する方針としてはどうか。</li> </ul>
真の病名	<ul style="list-style-type: none"> <li>電子カルテ情報には保険傷病名（保険請求目的）はあるが、「真の病名」がない。</li> </ul>	<ul style="list-style-type: none"> <li>ICD10の他に医師が考える「臨床病名」を電子カルテで入力できるようにすることが考えられ、実施している医療機関もある。</li> <li>米国では電子カルテからの表現型抽出（EHR Phenotyping）による「真の病名」あるいは記載されていない病名の推測方法について研究されており（※）、大規模DBへの適用も考えられる。  <small>※Electronic health record phenotyping improves detection and screening of type 2 diabetes in the general United States population: A cross-sectional, unselected, retrospective study. J Biomed Inform. 2016 Apr;60:162-8</small></li> </ul>
その他	課題	解決策/改善策の例
標準コードのバージョン更新及びデータ蓄積	<ul style="list-style-type: none"> <li>癌ステージ分類UCCなどは版が変わるが、蓄積されているデータには、その情報は通常は反映されていない。</li> </ul>	<ul style="list-style-type: none"> <li>ICD10からICD11、JLAC10からJLAC11のように大きな改訂では、変わるコード体系の名称自体が変わり、新旧が混在することはない。</li> <li>途中で、マイナーに変更があり、版が変わるコード体系については、コード体系とバージョンをあわせて記録する。</li> </ul>
施設ごとのオーダー運用等の違い	<ul style="list-style-type: none"> <li>施設により薬剤や検査のオーダーの運用の違いがあり、医療等情報の二次利用に影響する可能性がある。（例：COVID-19でワクチンが注射オーダーでなく処置オーダーとなっている場合がある。PCR検査結果が検査オーダーの場合と細菌検査オーダーの場合もある。）</li> <li>処置オーダーと処方オーダーで機能が異なり、データ内容も異なる場合もある。</li> <li>病院の運用は簡単には変更できない。</li> <li>ベンダー依存とは限らずローカルルールで動いている場合もある。</li> </ul>	<ul style="list-style-type: none"> <li>困難な課題ではあるが、RWDを生成する病院情報システムの機能として、医療等情報の二次利用の観点から課題を整理し、推奨を公表してはどうか。</li> </ul>

# 体外診断用医薬品の添付文書でのJLACコードの記載について①

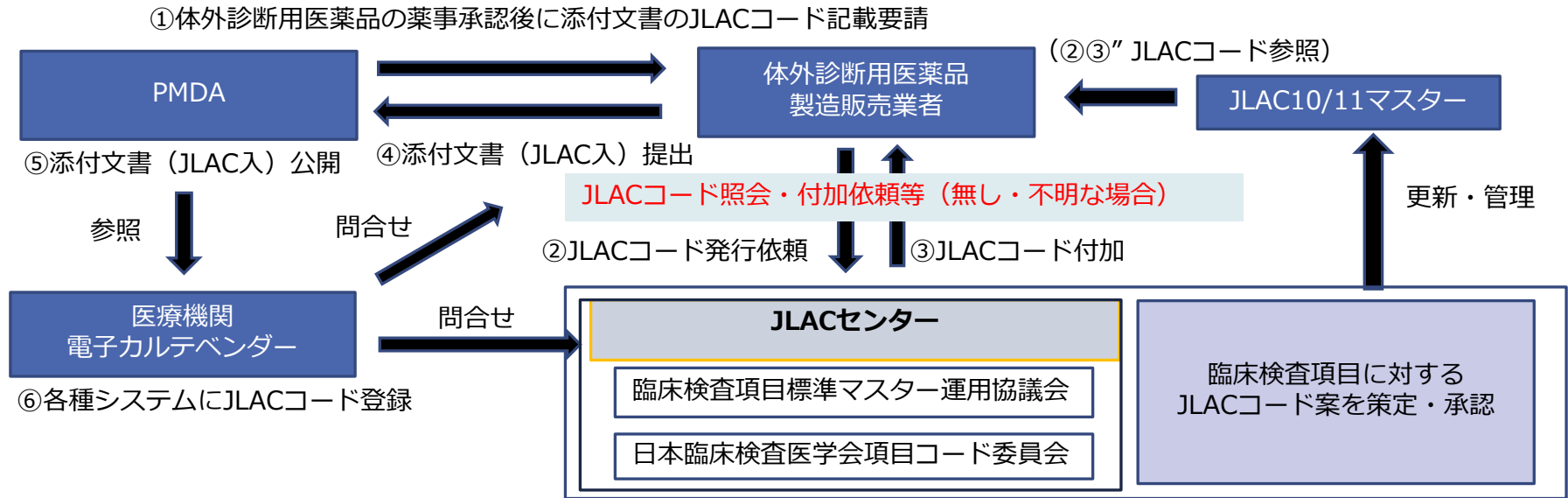
## 体外診断用医薬品の添付文書にJLACコード（臨床検査項目コード Japan Laboratory Code）の記載を要件にする場合の論点（案）

JLAC10コード	<ul style="list-style-type: none"> <li>● 17桁コード：[JLAC 検査名称：分析物（5桁）＋識別（4桁）＋材料（3桁）]＋測定法（3桁）＋結果識別（2桁）（特徴）</li> <li>● 1製品：複数のコードを付番することになる。</li> <li>● JLAC10コードは利用者が、JLACコードの5要素についてそれぞれ適切なコードをコード表全体から選択する方法であり、同じ検査でもコードの選択の違いにより17桁コードにバラつきが生じる。</li> <li>● 測定法の分類が細分化され、医療専門職（臨床検査技師等）も最適なコードを選択することが困難である。</li> </ul>
JLAC11コード	<ul style="list-style-type: none"> <li>● 17桁コード：[JLAC 検査名称：測定物（5桁）＋識別（4桁）＋材料（3桁）]＋測定法（3桁）＋単位（2桁）（特徴）</li> <li>● 1製品：依頼（オーダー）ではJLAC11コードが原則1：1となる。（※結果項目単位では複数コードとなるものもある）</li> <li>● JLAC11コードは、使用する体外診断用医薬品毎に測定法コードが付与され、他の4要素の適切なコードが絞り込まれている。</li> <li>● 識別、材料、結果単位を選択肢から選択する方法であり、JLAC10コードよりも簡易にマッピングできる。</li> </ul>
コーディング（付番）	「JLACセンター」が臨床検査項目にJLACコード案を策定し、日本臨床検査医学会項目コード委員会で承認する。
マッピング（採番）	各施設で、自施設実施の検査に合うJLACコードをコード表から採番し、施設内コードとJLACコードを連結させる。
必要性	<ul style="list-style-type: none"> <li>● 電子カルテ等の医療等情報のデータ交換規格として採用されるHL7 FHIRでは、データ項目に対して用いる標準コード体系とコードを割り当てる仕様であり、国内における臨床検査項目の標準コード体系とコードは「臨床検査項目分類コード（JLACコード）」がある。</li> <li>● 医療機関等はHL7 FHIRに準拠した3文書（①診療情報提供書、②退院時サマリー、③健康診断結果報告書）6情報（①傷病名、②アレルギー情報、③感染症情報、④薬剤禁忌情報、⑤検査情報、⑥処方情報）をFHIR出力を行う必要があるが、<b>検査情報</b>では医療機関等にJLACが導入されておらず、JLACを割当てている一部施設でもその正確性が検証されておらず、施設間で割当てられた標準コードが異なる課題がある。また、医療機関等から得られるHL7 FHIRの記述情報もバリデーションが担保されておらず、標準コードの適切性を検証する方法もない。</li> </ul> <p>→ <b>体外診断用医薬品の添付文書にJLAC11コードが正しく記載されることにより、病院情報システムの上流工程において、正しいJLAC11コードを院内に取り込むことを容易にできるのではないか。</b></p>
JLACコードの正確性確保	<ul style="list-style-type: none"> <li>● 一定期間の後に体外診断用医薬品の添付文書にJLACコードが記載される場合、JLACコードの正確性を諮問機関（臨床検査医学会・項目コード委員会）での承認を通じ、JLACセンターで確保する。</li> </ul>

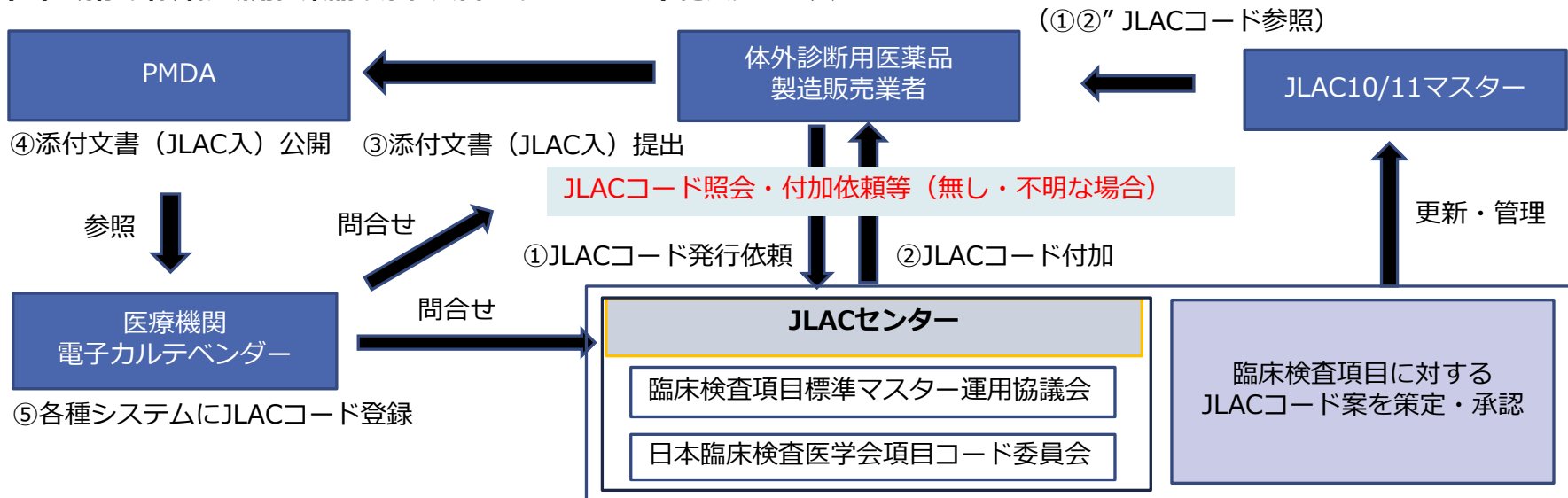


# 体外診断用医薬品の添付文書でのJLACコードの記載について②

## (1) 新規の体外診断用医薬品の添付文書へのJLACコード記入プロセス



## (2) 既存の体外診断用医薬品の添付文書へのJLACコード記入プロセス



# JLAC10コードとJLAC11コード

17桁

第1要素(5桁) 第2要素(4桁) 第3要素(3桁) 第4要素(3桁) 第5要素(2桁)

JLAC10	分析物コード	識別コード	材料コード	測定法コード	結果識別コード
JLAC11	測定物コード	識別コード	材料コード	測定法コード	結果単位コード

JLAC10の「識別コード」と「結果識別コード」の機能を組み合わせて構成

JLAC10をもとにした、より詳細な分類

JLAC10は測定原理で分類  
JLAC11は体外診断用医薬品の場合「測定法名称と商品名称の組み合わせ」で特定

JLAC10は結果の分類(定性, 定量, 等)  
JLAC11は結果の値の単位

## JLAC10とJLAC11は異なるコード体系 - 同じ要素名でも異なる分類体系

例) 総蛋白(TP)

### JLAC10コード

分析物	識別	材料	測定法	結果識別
3A010	0000	023	271	01
総蛋白		血清	可視吸光光度法	定量値

### JLAC11コード

測定物	識別	材料	測定法	結果単位
C1002	0401	250	027	55
蛋白	測定値	血清	可視吸光光度法「セロテック」TP-L	g/dL

1対1の対応関係ではない

## COVID-19 抗原検査キット「INNOVITA」 JLAC10コード

分析物	識別	材料	測定法	結果識別
5F625	1410	063	190	11
SARS-CoV-2	ウイルス抗原	鼻汁, 鼻腔・鼻咽頭擦過物	イムノクロマトグラフィ法	判定

COVID-19 抗原検査キット「INNOVITA」の添付文書  
[457008\\_30500EZx00027000\\_A\\_01\\_01.pdf \(pmda.go.jp\)](https://www.pmda.go.jp/457008_30500EZx00027000_A_01_01.pdf)  
 JLAC10コード表  
<https://www.jslm.org/committees/code/>

## COVID-19 抗原検査キット「INNOVITA」 JLAC11コード

測定物	識別	材料	測定法	結果単位
V2341	0403	345	080	02
SARS-CoV-2 抗原	判定	鼻腔・鼻咽頭の粘液・擦過物	イムノクロマトグラフィ法_COVID-19抗原検査キット「INNOVITA」	定性・判定

JLAC11コード表 (日本臨床検査医学会ホームページ)  
<https://www.jslm.org/committees/code/>

- JLAC11コード付番は、JLAC10コード付番と同様、添付文書の情報からJLAC11の各要素をコーディングする。
- JLAC11コードの「測定法」は、体外診断用医薬品毎に固有のコードが付番される。(1:1)

# COVID-19 抗原検査キット「INNOVITA」のJLAC10コード付番方法

## 添付文書の記載内容をもとにJLAC10コード表から適切なコードを選択する

分析物	識別	材料	測定法	結果識別
5 F625	1410	063	190	11
SARS-CoV-2	ウイルス抗原	鼻汁、鼻腔・鼻咽喉擦過物	イムノクロマトグラフィ法	判定

- ①分析物  
分析物コード：5 F625（分析物名：SARS-CoV-2）
- ②識別  
識別コード：1410（識別名：ウイルス抗原）
- ③材料  
材料コード：063（材料名：鼻汁，鼻腔・鼻咽喉擦過物）
- ④測定法  
測定法コード：190（測定法名：イムノクロマトグラフィ法）  
添付文書にはイムノクロマトグラフィ法の記載はないが、  
マーカー部の記載からJLAC10コード表の測定法コード190と判断
- ⑤結果識別  
結果識別コード：11（結果識別名：判定）

添付文書

COVID-19 抗原検査キット「INNOVITA」の添付文書  
[457008\\_30500EZ00027000\\_A\\_01\\_01.pdf \(pmda.go.jp\)](https://www.pmda.go.jp/457008_30500EZ00027000_A_01_01.pdf)

③材料

①分析物 ②識別

④測定法

⑤結果識別

### JLAC10コード表

JLAC10コード表  
<https://www.jslm.org/committees/code/>

#### ①分析物

分析物コード	分析物名(1)	分析物名(2)	
5F600	レオウイルス		reovirus
5F610	ロタウイルス		rotavirus
5F620	SARS コロナウイルス		SARS-CoV
5F624	SARS-CoV-2+インフルエンザA・B+RSウイルス		SARS-CoV-2
5F625	SARS-CoV-2		SARS-CoV-2
5F626	SARS-CoV-2+インフルエンザA・B		SARS-CoV-2

#### ②識別

コード	識別名	識別名(2)	識別名(英名)
1405	ウイルス遺伝子型		viral genotype
1406	ウイルス血清型		viral serotype
1410	ウイルス抗原		viral antigen
1411	ウイルス抗原定性		viral antigen, qualitative

#### ③材料

コード	材料名	Materials
060	分泌液	[Secretion]
061	精液	semen
062	喀痰	sputum
062	乳汁	milk
063	鼻汁、鼻腔・鼻咽喉擦過物	nasal drip, nasal cavity, nasopharynx exfoliative

#### ④測定法

コード	測定法名	測定法名(2)	
●その他の免疫学的測定法			
181	酵素免疫固定法		IAA(In)
186	免疫不活化法		不活化
190	イムノクロマトグラフィ法		

#### ⑤結果識別

コード	結果識別名	
●区分表現		
11	判定	判定
12	反応	反応

# 參考資料

# データの品質保証(Data Quality Assurance)の議論

## ベリフィケーション(verification)とバリデーション(validation)

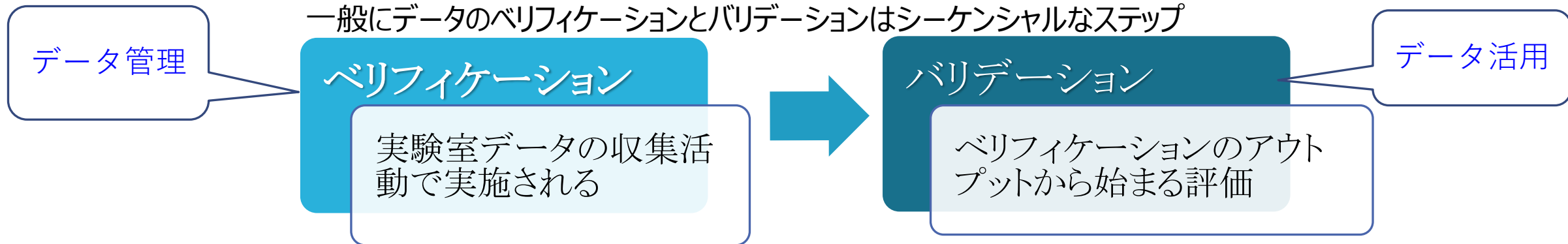
環境データ・ベリフィケーション及びデータ・バリデーションに関するガイダンス (米国環境保護庁, US Environmental Protection Agency)

### データ・ベリフィケーション：

- データセットの完全性、正確性、適合性 (conformance/compliance)を、方法および手続き的要件、契約要求事項に照らして評価するプロセス
- 分析的手法の手続きや契約要求で事前に定義された(しばしばgenericな)要件に照らしての評価

### データ・バリデーション：

- 検査対象(analyte-)およびサンプル固有のプロセスで、データ・ベリフィケーションを超えて特定のデータセットの分析的クオリティを評価すること
- プロジェクト固有の文書における「プロジェクトに特有のデータニーズ」に焦点を当てている



※ほとんど常にデータ収集者やデータ利用者とは独立のパーティにより実施される

# FDA RWEのフレームワーク

## レギュラトリ・ディシジョンの利用におけるRWD適合性評価(Assessing Fitness of RWD)

**RWEの強度:** 臨床試験の方法とData Reliabilityおよび根底にあるData Relevanceに依存

※ Sentinel SystemではRWEの生成のためのRWD評価にData ReliabilityとRelevanceを検討

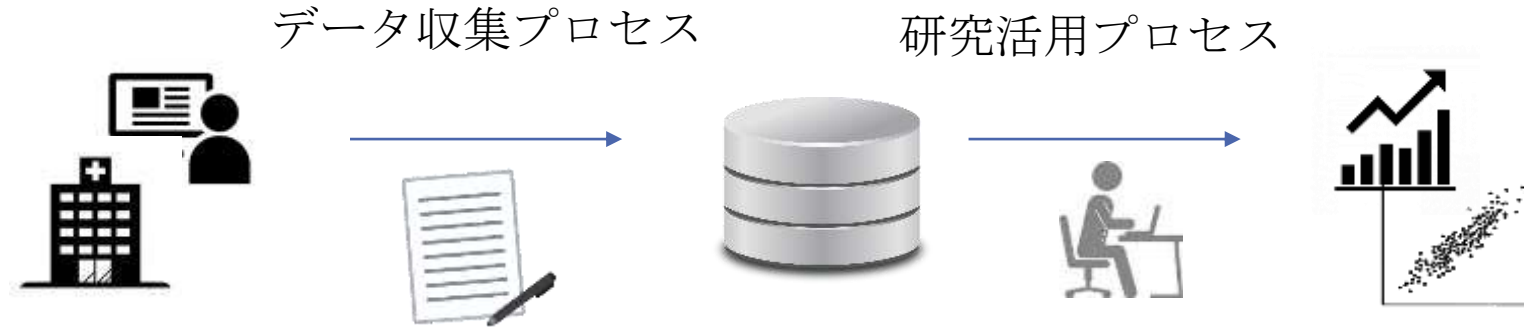
データの信頼性(Reliability)	データの適切性(Relevancy)
データ収集プロセス( <u>accrual</u> )と データ品質管理/品質保証( <u>quality control/assurance</u> )	リサーチ・クエスチョンごと
<b>Data Accrual:</b> RWDのソースはどうデータ要素が収集されるか、データ要素の定義、データ集積・統合の方法、適切な時間窓(relevant time windows)などが予め定められていること	<b>Dataのfit-for-purpose :</b> Fit-for-purposeは利用に照らしての適合性 この議論は複雑
<b>Reliability Assessment</b> コードやコードの組み合わせが適切に根底にある意図する医療コンセプトを表現できているか	<b>Relevance Assessment</b> データはfit-for-purpose であるか、暴露、アウトカム、共変量に関するデータがキャプチャされているかのアセスメントを含む

- 安全性評価に関してEHRのRWDを用いるにあたっての企業とスタッフへのガイダンス発行

*Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data*

# FDA RWEのフレームワーク

## レギュラトリ・ディシジョンの利用におけるRWD適合性評価(Assessing Fitness of RWD)



### Reliability Assessment

コードやコードの組み合わせが適切に根底にある意図する医療コンセプトを表現できているか

### Relevance Assessment

データはfit-for-purposeであるか、暴露、アウトカム、共変量に関するデータがキャプチャされているかのアセスメントを含む

データ品質管理/品質保証(quality control/assurance) :

品質保証の原則については国際的に合意された一定の基準があるわけではない

# Determining Real-World Data's Fitness for Use and the Role of Reliability, Duke Margolis Center for Health Policy, September 26, 2019

Figure 2. Examples of Quality Check Principles

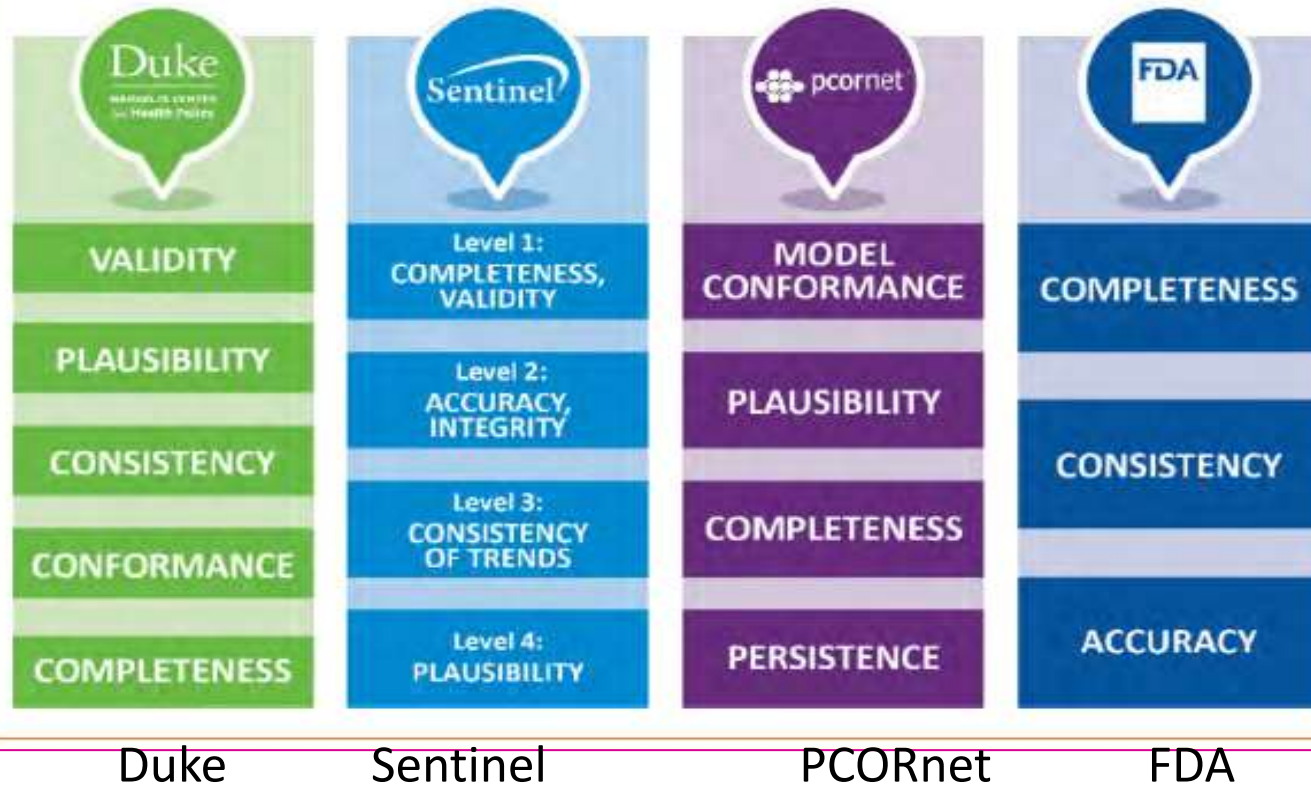
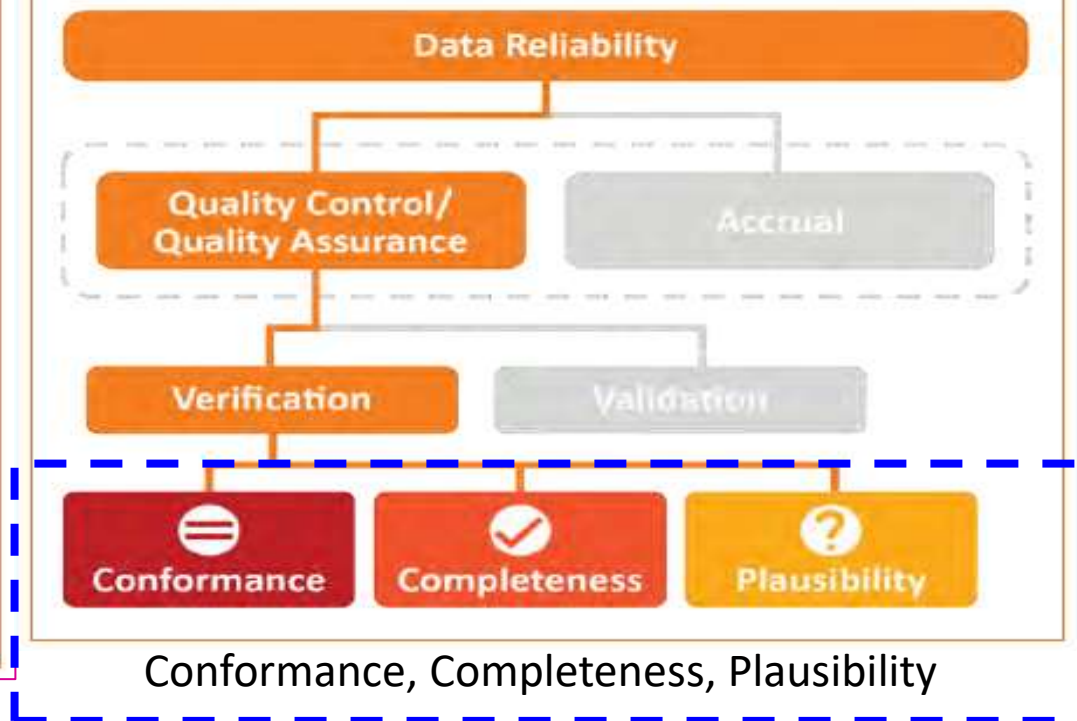


Figure 3. How Verification Checks for Completeness, Conformance, and Plausibility Contribute to Assessing Reliability to Support Fit-for-Use RWD\*



NOTE: FDA has identified issuing guidance on assessments of the reliability and relevance of RWD in generating RWE on drug product effectiveness as a program item in its 2018 RWE Framework. This paper does not reflect the views, guidance, or recommendations of FDA.



# データ品質アセスメント(Data Quality Assessment: DQA)

Kahn等はBiomedical Informaticsにおけるデータ品質評価に関する既存のターミノロジーを単一化させ(unify)、調和されたDQAターミノロジーを整理・提案

Kahn M, Callahan T, Barnard J, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. EGEMS (Wash DC). 2016;4(1):1244. doi: 10.13063/2327-9214.1244. eCollection 2016

調和されたDQAターミノロジー  
3つのカテゴリ

- 適合性(Conformance)はデータの相互利用可能性(Interoperability)に強く関連
- 完全性(Completeness)はRCTで前向きに収集されたデータに比べてRWDは低い

## データ標準とデータ品質

- データ・ベリフィケーションにおけるデータ品質、特に適合性(Conformance)の観点においてデータ相互運用性のためのデータ標準の重要性
- データ・バリデーションのコンテキストからのデータ品質チェックは重要であるが、データ・ベリフィケーションのコンテキストにおけるチェックに比べ、定義、実施とも複雑

## Conformance (適合性)

データが以下に適合する度合い：

- ① あらかじめ定義された内的または外的なformatting制約への適合 (値適合: Value Conformance)
- ② 構造的制約への適合
- ③ 既存の値に基づいて計算して導出された値の正確性 (計算適合: Calculation Conformance)

## Completeness (完全性)

データの存在が以下の2つの観点からアセスされること：

- ① 時間軸の一点 (Atemporal Completeness)
- ② 複数の時点 (Temporal Completeness)

## Plausibility (質が合理的で真であるらしい度合い)

The quality of being reasonable and likely to be true :

信じるに足る程度が次のように評価(assess)されること：

- ① 論理的に制約された測定や分布、同一の事実についての独立な複数の測定が一致すること (Atemporal Plausibility)
- ② 時間軸の属性や順序、測定値の状態遷移(Temporal Plausibility)
- ③ 重複する測定値、変数の値または記録の存在 (Uniqueness Plausibility)



第2回医療等情報の二次利用に関する技術作業班  
2024年3月6日、15:30-17:30

第2回医療等情報の二次利用に関する  
技術作業班（令和6年3月6日）  
資料2-2

## 多施設電子カルテデータの二次利用の取り組みから

大阪公立大学大学院医学研究科  
臨床検査・医療情報医学  
岡村 浩史

# 自己紹介

---

1. 工学部卒業、大学院修士課程修了
2. 民間企業でシステムエンジニア
3. 医学部 3 年次編入
4. 大阪市立大学で血液内科医として診療
5. 大阪市立大学で臨床医学博士課程
6. 血液内科診療 + 医療情報学

# 多施設電子カルテデータの二次利用

---

1. 全ゲノム解析等実行計画（AMED + 事業実施組織）
  - Genomics England
  - JASPEHRプロジェクト
2. 造血幹細胞移植患者の臨床判断支援システム開発（AMED）
  - 標準コードの付番・OMOP
  - 他の臨床試験との衝突リスク

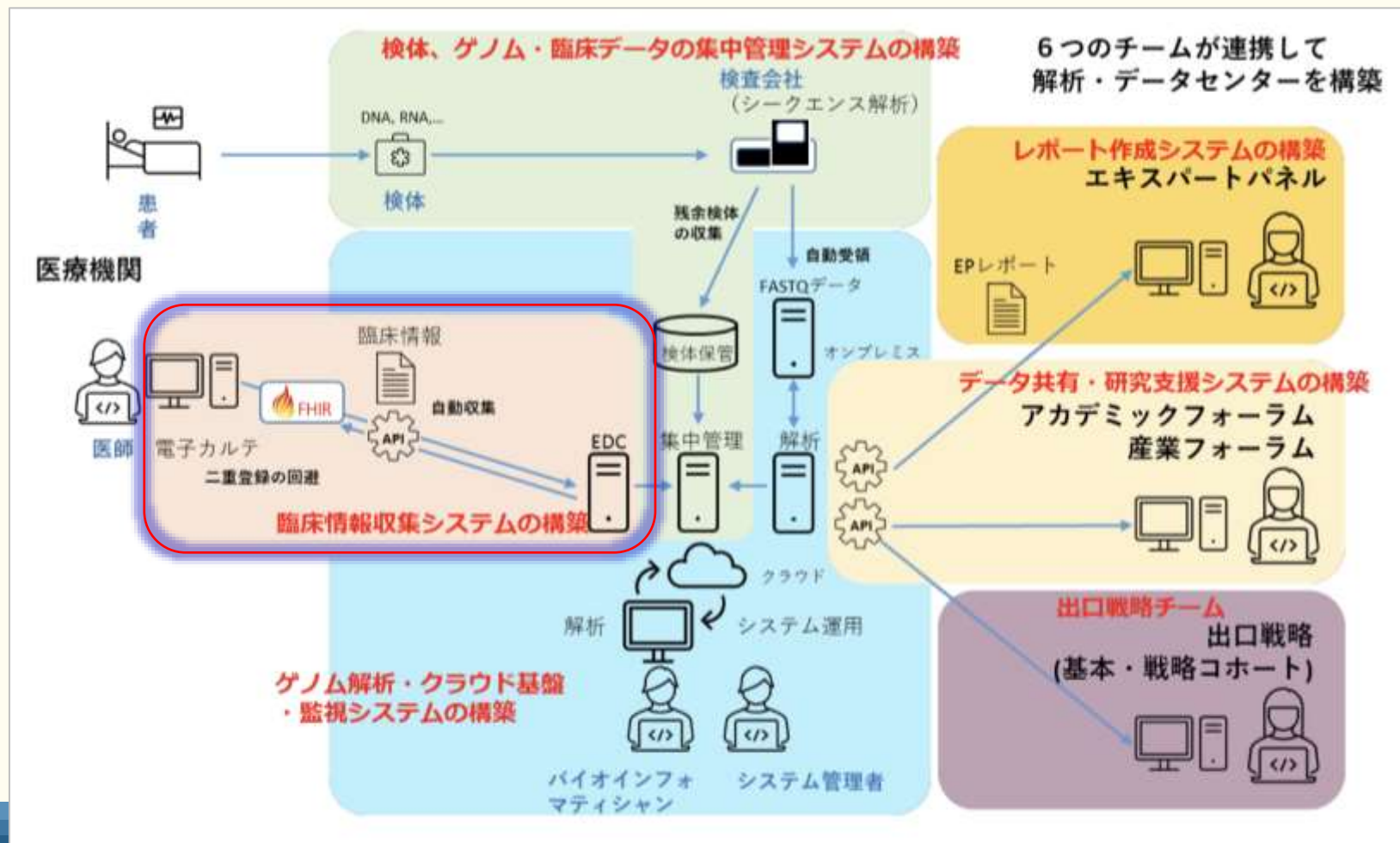
# 多施設電子カルテデータの二次利用

---

1. 全ゲノム解析等実行計画（AMED + 事業実施組織）
  - Genomics England
  - JASPEHRプロジェクト
2. 造血幹細胞移植患者の臨床判断支援システム開発（AMED）
  - 標準コードの付番・OMOP
  - 他の臨床試験との衝突リスク

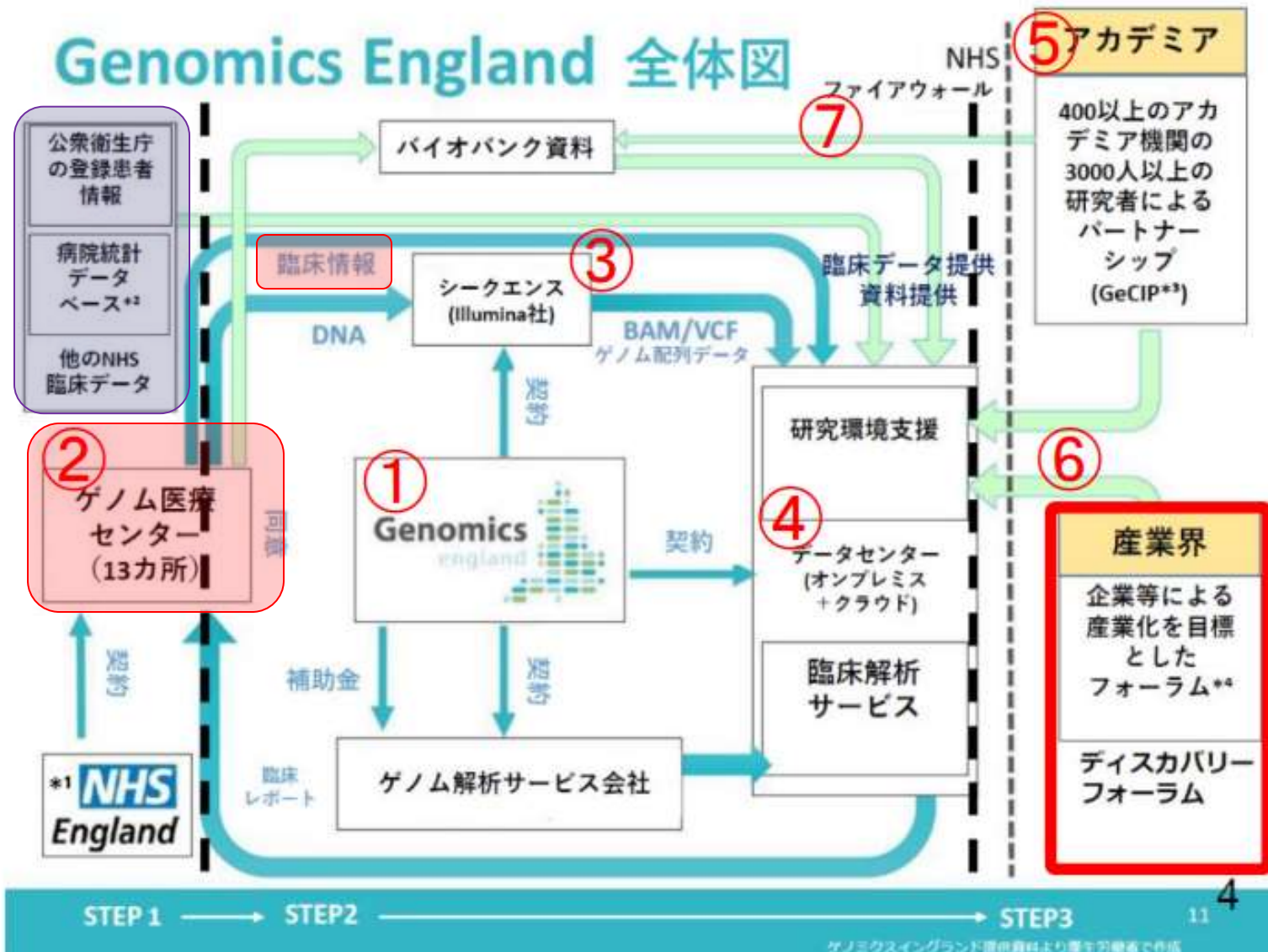
# 全ゲノム解析等実行計画

がん患者と難病患者の全ゲノムデータと診療情報を統合した情報活用基盤を構築する。



# Genomics England

# Genomics England





# Genomics England

---

## 100k Genomes Project NGRL data release v18 - December 2023

### Genomes\*

#### 106,263 genomes

- 32,751 Cancer (48.5% Germline)
- 73,512 Rare Disease

### Primary clinical data

#### 90,178 participants

- 17,241 Cancer
- 72,937 Rare Disease

### Secondary data

- Hospital Episode Statistics (HES)
- Mental Health (MHMD, MHLDDS, MHSDS)
- COVID test result data

- NCRAS - history of treatment data for patients diagnosed with Cancer between 1995 and December 2019
- Uncurated SACT (chemotherapy) data

### Clinically interpreted data & QC

- 34,020 families with Tier 1, 2 and 3 variants from interpretation pipeline
- 33,533 families with GMC exit questionnaires
- 62,383 tiered and quality checked rare disease genomes; 31,539 quality checked cancer genomes

\*Uniquely sequenced samples for consented participants of this release.

[https://re-docs.genomicsengland.co.uk/data\\_overview/](https://re-docs.genomicsengland.co.uk/data_overview/)

# Genomics England



研究者



- 患者背景
- 家族歴
- 疾患名
- Stage
- 画像検査レポート
- 手術、化学療法  
など



...



...

Secondary clinical data

Primary clinical data

# 全ゲノム解析等実行計画



研究者



Visiting環境

データベース

ゲノム情報

HIC

- 患者背景
- 家族歴
- 疾患名
- Stage
- 画像検査レポート
- 手術、化学療法  
など

NDB  
DPC  
6情報

行政機関  
死亡情報

全国がん登録

A施設

B施設

Secondary clinical data

Primary clinical data

# JASPEHRプロジェクト

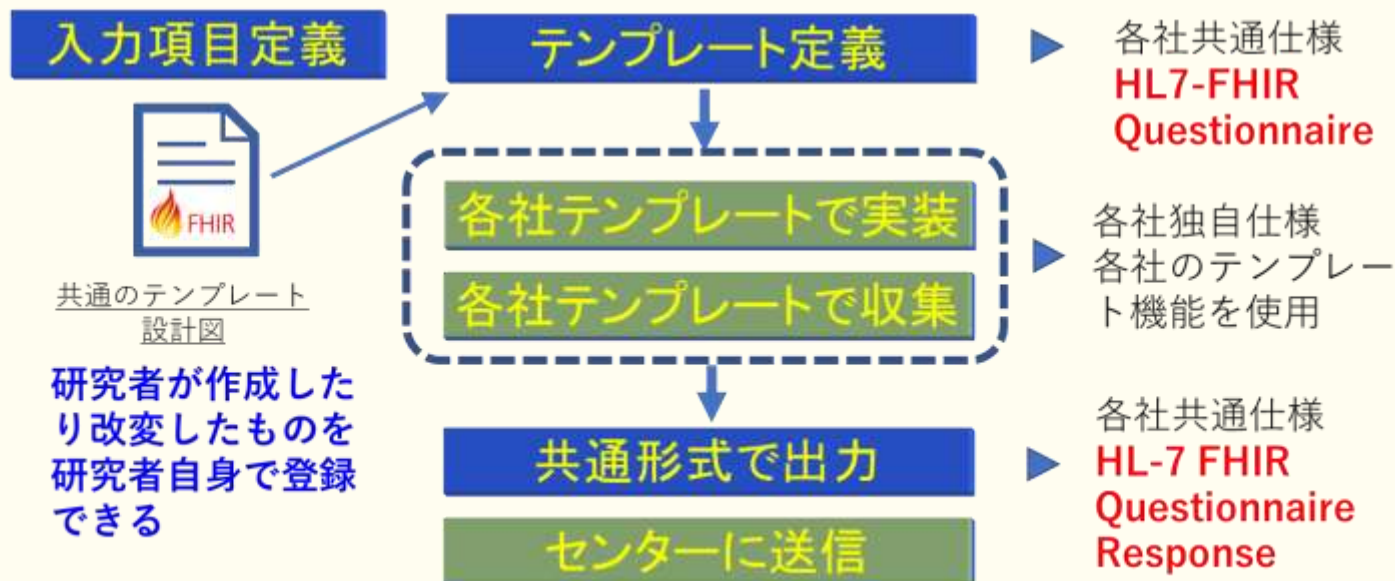
# JASPEHRプロジェクト

## 医療機関側の入力テンプレートはJASPEHR方式を導入

記述仕様には医療情報標準規格であるHL7-FHIRを採用

疾患ごとにテンプレートを開発するのではなく、  
様々なテンプレートに対応できる仕組みを開発する

JASPEHRプロジェクト (JApanese Standard Platform for EHRs)



[https://cmii.ncgm.go.jp/project\\_jaspehr.html](https://cmii.ncgm.go.jp/project_jaspehr.html)

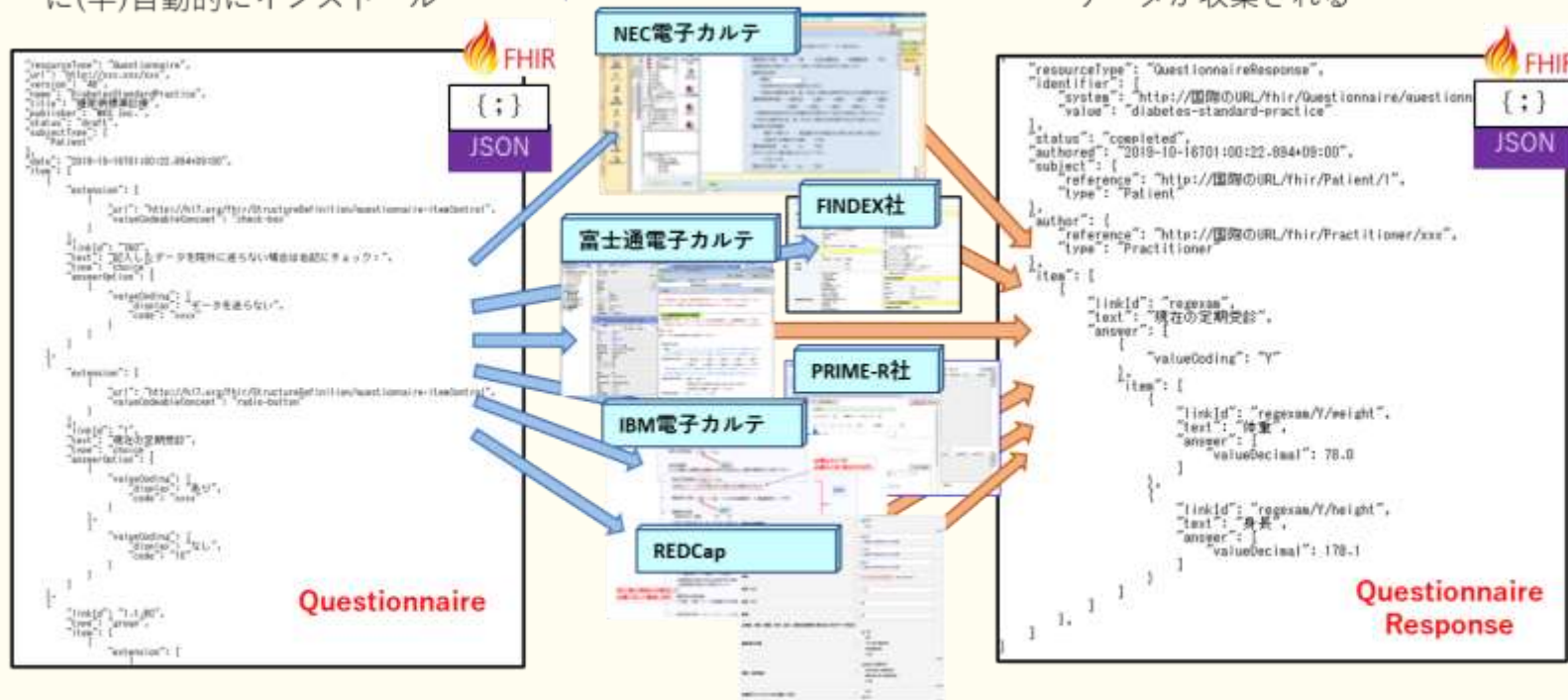
# JASPEHRプロジェクト



学会が定めた診療に必須の入力項目を各社の電子カルテに(半)自動的にインストール

診察時に、電子カルテにテンプレートを  
使って診療記録を記載

電子カルテが異なっても同一形式の診療  
データが収集される



システムの改修を追加で行うことなく、  
書式の作成・展開を行うことで、対応が可能

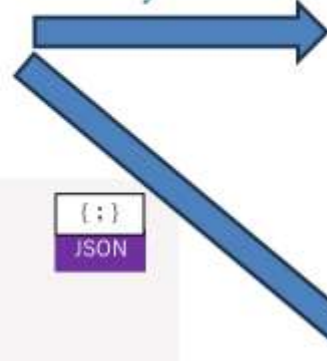
# JASPEHRプロジェクト



- Questionnaire\_WGS\_がん種情報 (固形腫瘍)\_20230315.json
- Questionnaire\_WGS\_がん種情報 (造血器腫瘍)\_20230315.json
- Questionnaire\_WGS\_全ゲノム解析後\_基本情報\_20230315.json
- Questionnaire\_WGS\_必要時 (同意変更情報)\_20230315.json
- Questionnaire\_WGS\_必要時 (症例管理情報)\_20230315.json
- Questionnaire\_WGS\_患者背景\_家族歴+非薬物療法\_WHO\_20230315.json
- Questionnaire\_WGS\_既往歴併存疾患\_20230315.json
- Questionnaire\_WGS\_検体情報 (固形腫瘍)\_20230315.json
- Questionnaire\_WGS\_検体情報
- Questionnaire\_WGS\_症列基
- Questionnaire\_WGS\_薬物療
- Questionnaire\_WGS\_転帰
- Questionnaire\_WGS\_造血幹
- Questionnaire\_全ゲノム解析
- Questionnaire\_全ゲノム解析
- Questionnaire\_全ゲノム解析
- Questionnaire\_検体データ
- Questionnaire\_遺伝性疾患コ

```

{
  "resourceType": "Questionnaire",
  "id": "123",
  "name": "HL7 FHIR Questionnaire",
  "type": "Form",
  "status": "draft",
  "subjectType": "Patient",
  "extension": [
    {
      "url": "http://hl7.org/fhir/StructureDefinition/questionnaire-checkControl",
      "valueBoolean": true
    }
  ],
  "item": [
    {
      "type": "group",
      "item": [
        {
          "type": "text",
          "label": "患者背景",
          "text": "患者背景"
        }
      ]
    }
  ]
}
    
```



初発/再発の別  初発  再発

Rai分類 (CLL)  stage 0  stage 1  stage 2  stage 3  stage 4

Binet分類 (CLL)  stage A  stage B  stage C  不明

Lugano分類 (消化管MALT)  stage I  stage II  stage IIE  stage IV  不明

Murphy分類 (小児Burkitt)  stage I  stage II  stage IIE  stage IV  不明

ATL/COG分類 (ATL)  慢性型 (予後不良因子なし)  慢性型 (予後不良因子あり)  急性型  不明

Ann Arbor分類 (その他リンパ腫)  stage I  stage II  stage III  stage IV  不明

Ann Arbor分類 (その他リンパ腫) B症状  A  B  不明

その他

HTLV-1

HIV  初発  再発

EBV (EBER)  stage 0  stage 1  stage 2  stage 3  stage 4  不明

ENL201  stage A  stage B  stage C  不明

Favoral  stage I  stage II  stage III  stage IV  不明

富士通電子カルテ

データの標準化には、現場の運用の標準化が必要条件となることが多々ある！

- 診療の業務（＝現場の運用）を標準化し、統一形式でデータ収集ができる。  
⇒ あらゆる疾患領域に応用可能
- レジメン、クリニカルパス、臨床研究プロトコルに応用すれば、それらの運用も標準化し、統一形式でデータ収集ができる。  
  
⇒ 施設やベンダー間の垣根を越え、様々な診療業務の運用と情報を共に標準化することができ、
  - ・ 診療の標準化
  - ・ データの標準化の一石二鳥が実現し得る。

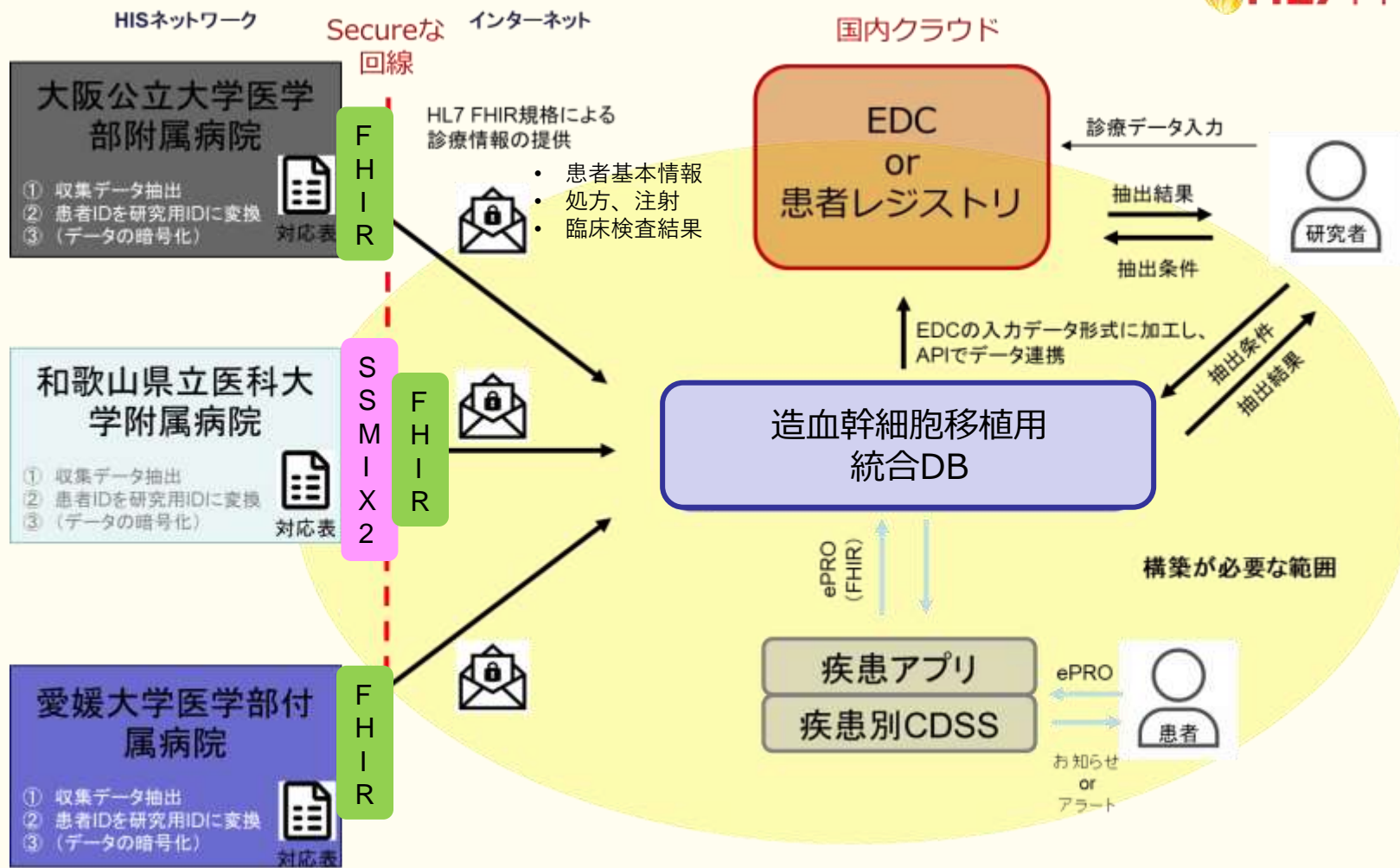


# 多施設電子カルテデータの二次利用

---

1. 全ゲノム解析等実行計画（AMED + 事業実施準備室）
  - Genomics England
  - JASPEHRプロジェクト
2. 造血幹細胞移植患者の臨床判断支援システム開発（AMED）
  - 標準コードの付番・OMOP
  - 他の臨床試験との衝突リスク

# 造血幹細胞移植患者の臨床判断支援システム開発



# 標準コードの付番・OMOP

# 標準コードの付番

---

JLAC10 or 11の普及は進んでいない。

- ✓ J-DREAMS
- ✓ J-CKD-DB
- ✓ . . .
- ✓ 電子カルテ情報共有サービス(6情報)

プロジェクト毎に、収集する項目に標準コード付番を各医療機関に求めている。

= 部分最適

本邦の全体最適として、誰がどのタイミングでどの範囲を付番するのが適切なのか。

⇒ ワークフローの整備 (JLACセンター)

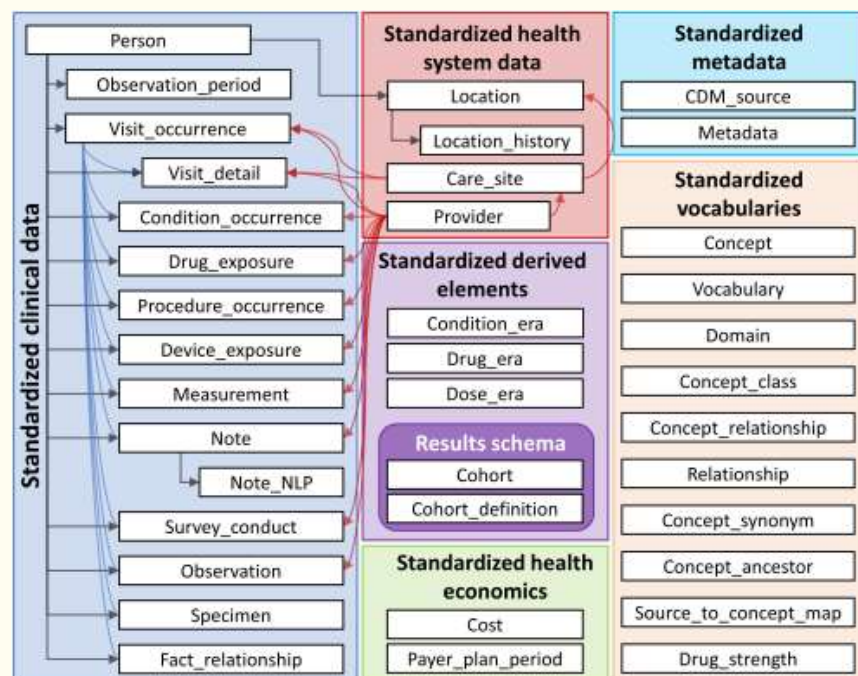
## 標準的医療情報システムに関する検討会（内閣官房健康・医療戦略室）報告書（令和元年11月29日）

- ハウスコードの標準コードへの変換が必要。
- 検査コードについて、試薬の承認時等にJLACコード対応を行うべき。

## 実装に向けた課題

- JLAC10コードは、2020年11月11日時点で6,711コード。
- 電子カルテ導入の一般病院のうち、検査項目にJLAC10コードの対応付け（マッピング）を実施している施設は24%。
- 実施中の施設のうち、JLACマッピング作業に必要な人材や予算等を確保できている施設は26%。

# OMOP



- 8.1億人の患者臨床データを保持
- 331個のデータベース：284個のEHR、28個のレセプト情報を含む
- 共通データモデルのため、複数DBに対して共通の解析プログラムが適用でき、実際、解析を支援する多くの分析ツールが提供されている。
- 同一の解析プログラムを各DBに配布し、各DBの分析結果を集めて統合する  
Network study方式では、各施設の個票を外部に出す必要がない。

# OMOPを用いた大規模臨床研究

## Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis

Marc A Suchard<sup>1</sup>, Martijn J Schuemie<sup>1</sup>, Harlan M Krumholz<sup>1</sup>, Seng Chan You<sup>1</sup>, Ruijun Chen<sup>1</sup>, Nicole Pratt<sup>1</sup>, Christian G Reich<sup>1</sup>, Jon Duke<sup>1</sup>, David Madigan<sup>1</sup>, George Hripcsak<sup>1</sup>, Patrick B Ryan<sup>1</sup>

Lancet. 2019;394:1816-1826.

- USA、韓国、日本、ドイツ
- **OMOP**を用いて、**490万人の高血圧患者**における複数の降圧薬の臨床転帰に対する効果と安全性を包括的に比較

## Characteristics and Outcomes of Over 300,000 Patients with COVID-19 and History of Cancer in the United States and Spain

Elena Roel<sup>1,2</sup>, Andrea Pistillo<sup>1</sup>, Martina Recalde<sup>1,2</sup>, Anthony G. Sena<sup>3,4</sup>, Sergio Fernández-Bertolín<sup>1</sup>, Maria Aragón<sup>1</sup>, Diana Puente<sup>1,2</sup>, Waheed-Ul-Rahman Ahmed<sup>5,6</sup>, Heba Alghoul<sup>7</sup>, Osaid Alser<sup>8</sup>, Thamir M. Alshammari<sup>9</sup>, Carlos Areia<sup>10</sup>, Clair Blacketer<sup>3</sup>, William Carter<sup>11</sup>, Paula Casajust<sup>12</sup>, Aedin C. Culhane<sup>13,14</sup>, Dalia Dawoud<sup>15</sup>, Frank DeFalco<sup>3</sup>, Scott L. DuVall<sup>16,17</sup>, Thomas Falconer<sup>18,19</sup>, Asieh Golozar<sup>20,21</sup>, Mengchun Gong<sup>22</sup>, Laura Hester<sup>23</sup>, George Hripcsak<sup>18,19</sup>, Eng Hooi Tan<sup>24</sup>, Hokyun Jeon<sup>25</sup>, Jitendra Jonnagaddala<sup>26</sup>, Lana Y.H. Lai<sup>27</sup>, Kristine E. Lynch<sup>16,17</sup>, Michael E. Matheny<sup>28,29</sup>, Daniel R. Morales<sup>30,31</sup>, Karthik Natarajan<sup>18,19</sup>, Fredrik Nyberg<sup>32</sup>, Anna Ostropolets<sup>18</sup>, José D. Posada<sup>33</sup>, Albert Prats-Urbe<sup>24</sup>, Christian G. Reich<sup>34</sup>, Donna R. Rivera<sup>35</sup>, Lisa M. Schilling<sup>11</sup>, Isabelle Soerjomataram<sup>36</sup>, Karishma Shah<sup>5</sup>, Nigam H. Shah<sup>33</sup>, Yang Shen<sup>22</sup>, Matthew Spotniz<sup>18</sup>, Vignesh Subbian<sup>37</sup>, Marc A. Suchard<sup>38</sup>, Annalisa Trama<sup>39</sup>, Lin Zhang<sup>40,41</sup>, Ying Zhang<sup>22</sup>, Patrick B. Ryan<sup>3,18</sup>, Daniel Prieto-Alhambra<sup>24</sup>, Kristin Kostka<sup>34</sup>, and Taliita Duarte-Salles<sup>1</sup>

BMJ Open. 2021;11:e057632.

- USA、ドイツ、オランダ、フランス、イタリア、スペイン、イギリス、韓国、中国
- **OMOPのデータベースから、285万人のCOVID-19患者**において、高血圧とCOVID-19重症化の関連を検討

# Genomics England と OMOP

[Genomic Medicine](#)[Our Initiatives](#)[Patients and Participants](#)[Research and Partnerships](#)[News and Events](#)[About Us](#)

Blog > Expertise, experience, and endurance: mapping NHS cancer records to a common data model

## Expertise, experience, and endurance: mapping NHS cancer records to a common data model

By Laura Kerr, Abigail Carter, and Helen Carter on 5 Jan 2024

### Public OMOP Mappings from Genomics England

#### Description

This repository contains mappings of NHS data sets to the Observational Medical Outcomes Partnership (OMOP) common data model version 5.4. Each mapping is supplied in two formats - a human-readable PDF and a machine-readable tab-separated variable (TSV) file. The PDFs should be fairly intuitive, and documentation is supplied for the machine-readable files.

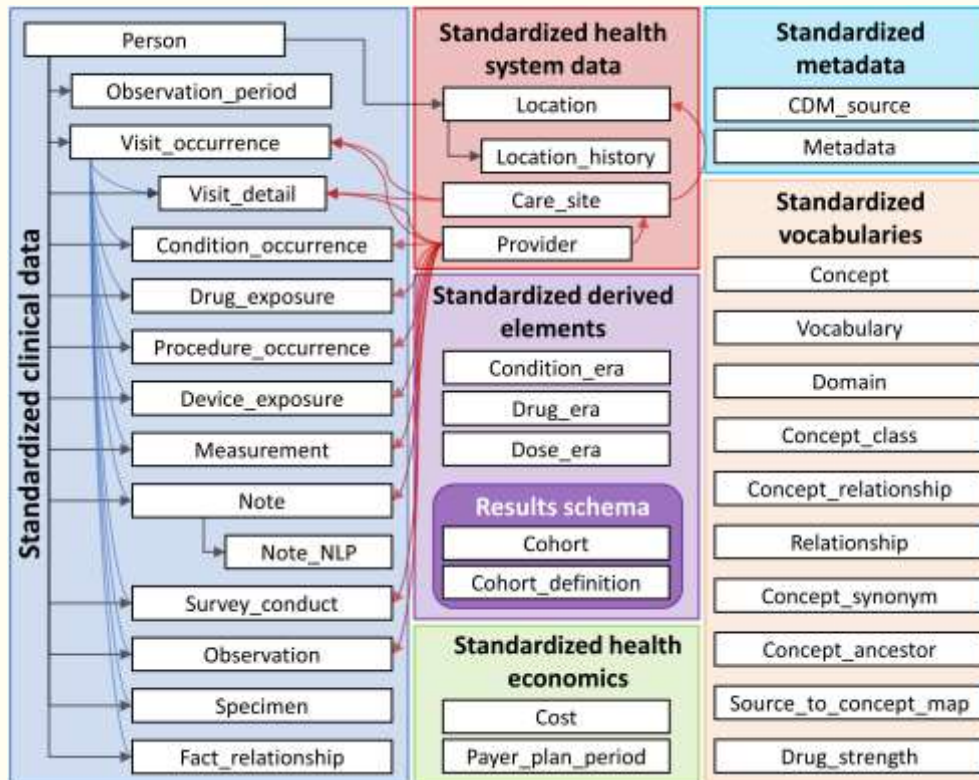
Where applicable, fact relationships have been defined. Most of these are "Has subject relationship context (SNOMED)" / "Subject relationship context of (SNOMED)" but there are some instances where more accurate representations of the relationships are possible, such as "Observation to Measurement" / "Measurement to Observation".

We have endeavoured to map to standard concepts wherever possible, using the SNOMED and LOINC vocabularies, but this has not been possible in all cases. Many of the NHS data sets contain enumerated values, which often have very vague meanings, and we have taken care not to add any information to the OMOP mappings that cannot be safely inferred. For this reason, some of the mappings use non-standard concepts and different vocabularies. Additionally, we have occasionally had to map a given source attribute to more than one OMOP concept and/or domain and link them via fact relationships.

In order to support traceability, we have defined some short identifiers for each data set to be used as source value prefixes. If these are employed, a source value of 'Consultant' in the Diagnostic Imaging data set would appear as "DID: Consultant" in the OMOP source value.

Genomics EnglandもOMOPマッピングを進めている。

# OMOP



- DXによって、臨床研究が国際化することは必然
- 標準コードは、国際標準コード（SNOMED、RxNorm、LOINCなど）へのマッピングが必要

⇒ 国際共同研究を見据えた二次利用の検討を



# 他の臨床試験との衝突リスク

# 他の臨床試験との衝突リスク

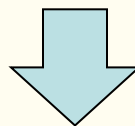
---

- DXによって、大規模なデータが収集され、リアルタイム性も高まる。  
⇒ 他の臨床試験・治験に参加している患者のデータが、二次利用される可能性

例)

疾患Aに対して適応拡大を目指す薬剤  $\alpha$  に関する治験に参加している患者情報が、二次利用データに含まれ研究利用できる状態

- ⇒ 他の研究者が、疾患Aに対する薬剤  $\alpha$  の効果を検討する後方視研究が可能
- ⇒ **施設と治験依頼者間の契約に違反するリスク**



- 各施設は、activeな臨床試験の被験者であるか否か判別できる情報を管理し、二次利用DBに提供する必要がある？
- or
- 治験依頼者が、治験毎に本邦でユニバーサルな被験者キーを管理し、二次利用DBに提供する必要がある？

# まとめ

---

- U.K.では、NHSが管理する複数の公的DBをSUS（本邦でのHIC）を通じて、Genomics Englandと共有することで、データの効率的な利活用を行っている。  
⇒ 「一定要件を満たすDB」と「HICの公的DB」を連携できる枠組みの検討
- JASPEHRは、運用とデータの両者を標準化したい場面で有用であり、このコンセプトは、レジメンやクリニカルパスへの応用が期待される。
- 標準コードは、誰がどの範囲をどのタイミングで付番し、各施設に反映させることが適切か、整備に向けたワークフローの具体化とその標準化が求められる。
- OMOPは、国際共同研究の共通データモデルとして普及が進んでいる。本邦の医療情報二次利用においても臨床研究の国際化を見据えた国際標準コードの付番戦略が求められる。
- 収集患者数と収集項目が増え、情報鮮度が保たれることで、データの利用可能性は広がる一方で、他の臨床試験との衝突に対応する必要がある。

## 第2回 医療等情報の二次利活用に関する技術作業班

### 二次利用環境における情報セキュリティ

独立行政法人情報処理推進機構  
デジタル改革推進部 田辺里美

# お話しする内容 = 安全管理措置に関する要件

## 医療等情報の二次利用に関する技術作業班の検討事項

データ連携の全体モデル		<ul style="list-style-type: none"> <li>● データ連携の全体モデルの作成</li> </ul>	<p>クラウド環境 (DaaS)</p> <p>インターネット      データ登録</p> <p>検索ポータルサイト 公的DB 民間DB</p> <p>医療等情報提供一覧表 利活用申請 (アクセス権付与)</p> <p>Visiting解析環境 (仮想デスクトップ) 統計解析処理</p>
DB	データ連携標準化	<ul style="list-style-type: none"> <li>● 各DBのデータ項目・標準コード・連結ID等の整理</li> <li>● 公的DB、民間DB等のアーキテクチャの整理</li> <li>● 標準コードの管理・付与に関する課題</li> <li>● データマスター (ID) の整理</li> </ul>	
医療機関等	データ信頼性確保	<ul style="list-style-type: none"> <li>● 医療機関等のデータ入力・提出に係る課題</li> <li>● 医療機関等に対する標準コード実装の推奨</li> </ul>	<p>インターネット      データ登録</p> <p>検索ポータルサイト 公的DB 民間DB</p> <p>医療等情報提供一覧表 利活用申請</p> <p>閉域環境      閉域環境</p> <p>Visiting解析環境 統計解析処理</p>
情報連携基盤	技術的要件	<ul style="list-style-type: none"> <li>● 情報基盤・解析基盤の技術的要件の整理</li> <li>● 情報連携基盤のアーキテクチャの策定</li> </ul>	
	接続要件	<ul style="list-style-type: none"> <li>● 公的DBとの接続要件</li> <li>● 民間DBとの接続要件 (及び資格認定)</li> <li>● 利活用者端末との接続要件 (及び資格認定)</li> </ul>	
	クラウド等インフラストラクチャー Visiting環境	<ul style="list-style-type: none"> <li>● 情報連携基盤から各DBへのアクセス環境の検討 (DaaS環境、VPN接続を介したVisiting環境)</li> <li>● Visiting環境及びクラウド環境 (DaaS) の要件</li> <li>● 利活用者のVisiting解析環境への接続要件</li> </ul>	
	ユーザビリティ	<ul style="list-style-type: none"> <li>● ポータル等のユーザビリティ要件</li> </ul>	
	安全管理措置に関する要件	<ul style="list-style-type: none"> <li>● 認証方法の要件 (二要素認証)</li> <li>● ログ記録・保存・管理の要件</li> <li>● 暗号化の要件 (IPsec等)</li> <li>● 安全管理措置 (保護措置) の整理</li> <li>● インシデント発生時の対応策の整理</li> </ul>	
情報提供一覧表	<ul style="list-style-type: none"> <li>● 医療等情報提供一覧表 (データ一覧) の整理</li> </ul>	<p>※DaaS (Data as a Service) : ネットワーク接続を介してクラウド上のデータストレージ、統合、処理、分析サービスを提供するデータ管理サービス</p> <p>※医療等情報: 健康・医療・介護等に関する患者情報を含む医療等情報を含むデータ全般を想定</p> <p>※医療機関等: 病院、一般診療所、歯科診療所、助産所、薬局、訪問看護ステーション、介護事業者、医療情報連携ネットワーク運営事業者等を想定</p> <p>※利活用者: 企業 (製薬メーカー、医療機器メーカー等)、研究機関 (大学等)、行政機関 (地方公共団体) 等を想定</p>	
審査 (※別途検討)	<ul style="list-style-type: none"> <li>● 利用者の申請、契約及びアクセス権付与の方法</li> <li>● 審査委員会の審査体制、審査項目、手続きの策定</li> <li>● 利活用申請書のフォーマット様式の策定</li> </ul>		

# 医療機関での情報システム利用の環境の変化

これまで>>

- 病院の中だけで完結（院内ネットワークのみ）
- 外部との接続が必要な場合は、専用の回線を利用
- サプライヤー（薬、配膳、その他備品等）とはFAX/電話でのやりとり



現在>>

- 様々な情報システムの利用が進み外部との連携も加速
- 従前の「外部接続なし」文化から「外部接続あり」環境を守る文化への切替えが追い付かない（体制面、費用面、マインドセット）



徳島県つるぎ町立半田病院  
大阪府立病院機構 大阪急性期・総合医療センター

⇒外部との接続装置の脆弱性を利用し侵入

# 情報システム利用環境変化と攻撃者や手法の変化

収益化  
組織化



## <クラウド利用（環境や接続先の多様化）>

- 様々な機器・ソフトウェア等の脆弱性、電子メールなど、多様な経路
- 情報の窃取（販売目的）、業務の停止  
⇒ **ゼロトラストセキュリティ**

自己顕示欲  
腕自慢



## <インターネット利用>

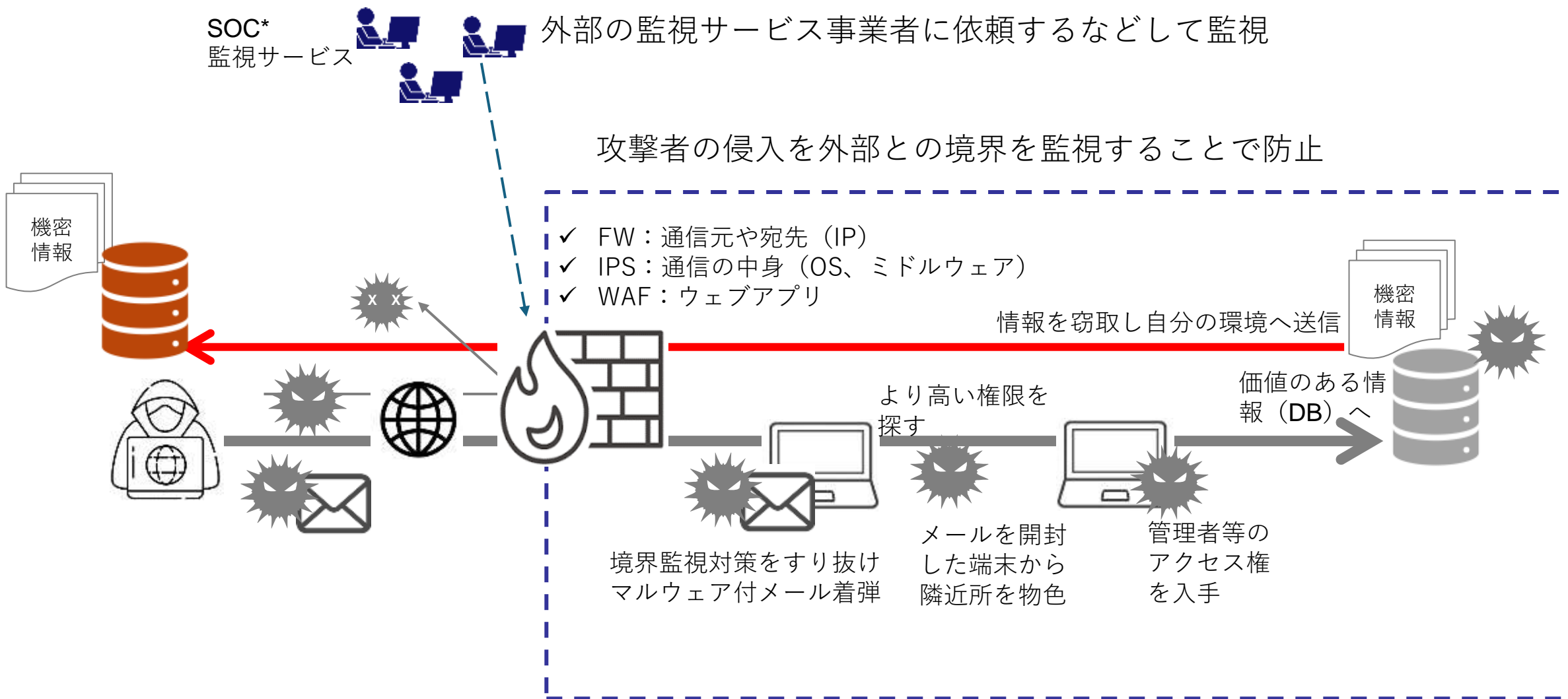
- 電子メールやホームページ経由
- コンピュータの性能低下・稼働停止による業務停止

知識追及  
興味

## <スタンドアロン>

- フロッピーディスク経由
- コンピュータの性能低下や破壊等の被害

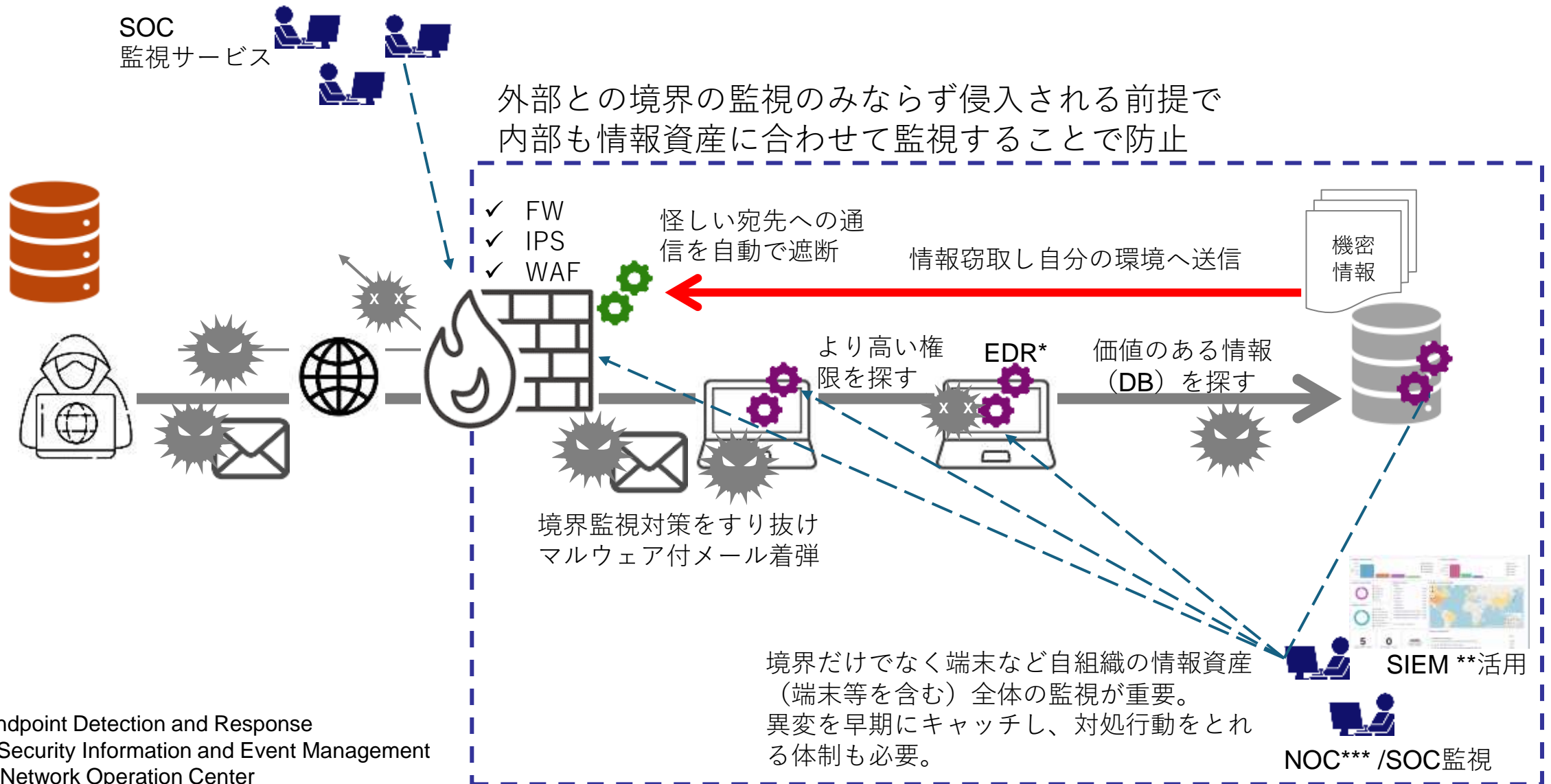
# 監視対象の変化（これまで：境界線の監視）



\*SOC: Security Operation Center



# 監視対象の変化（現在：利用者端末など情報資産ごとの監視も重要）

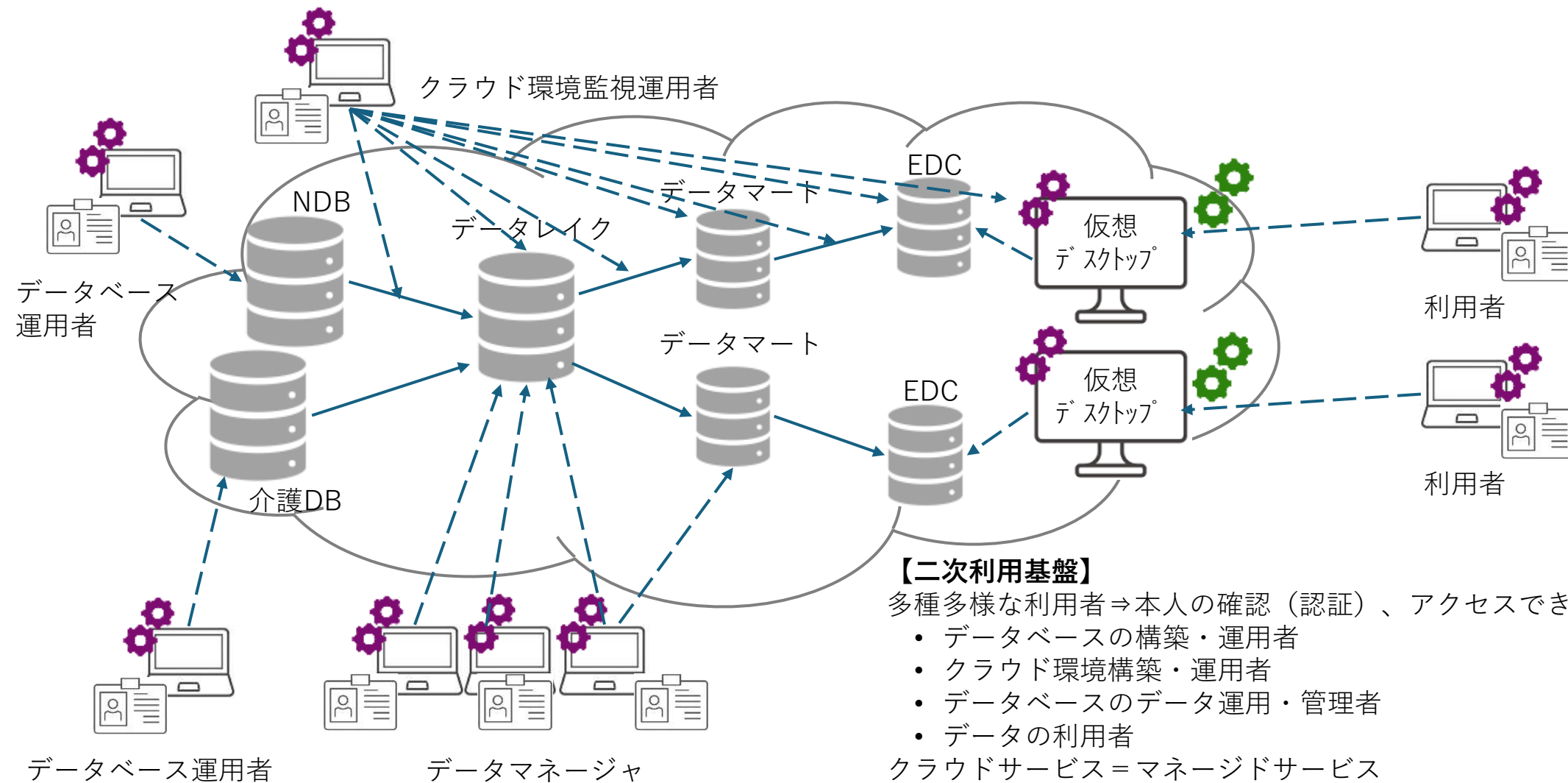


\*EDR: Endpoint Detection and Response

\*\*SIEM: Security Information and Event Management

\*\*\*NOC: Network Operation Center

# 「ゼロトラスト」の必要性 クラウド利用とセキュリティ



## 【二次利用基盤】

多種多様な利用者⇒本人の確認（認証）、アクセスできる範囲の確認（認可）

- データベースの構築・運用者
- クラウド環境構築・運用者
- データベースのデータ運用・管理者
- データの利用者

クラウドサービス＝マネージドサービス

- OSやアプリケーションなどのアップデート等は提供者が実施
- 環境の変化（アップデート）等を監視し設定不備を修正（ベスプラの活用）
- クラウド内部で発生する通信も監視し不審な挙動がないか監視が必要

# 二次利用基盤セキュリティ管理の概観

## 【ログの活用】

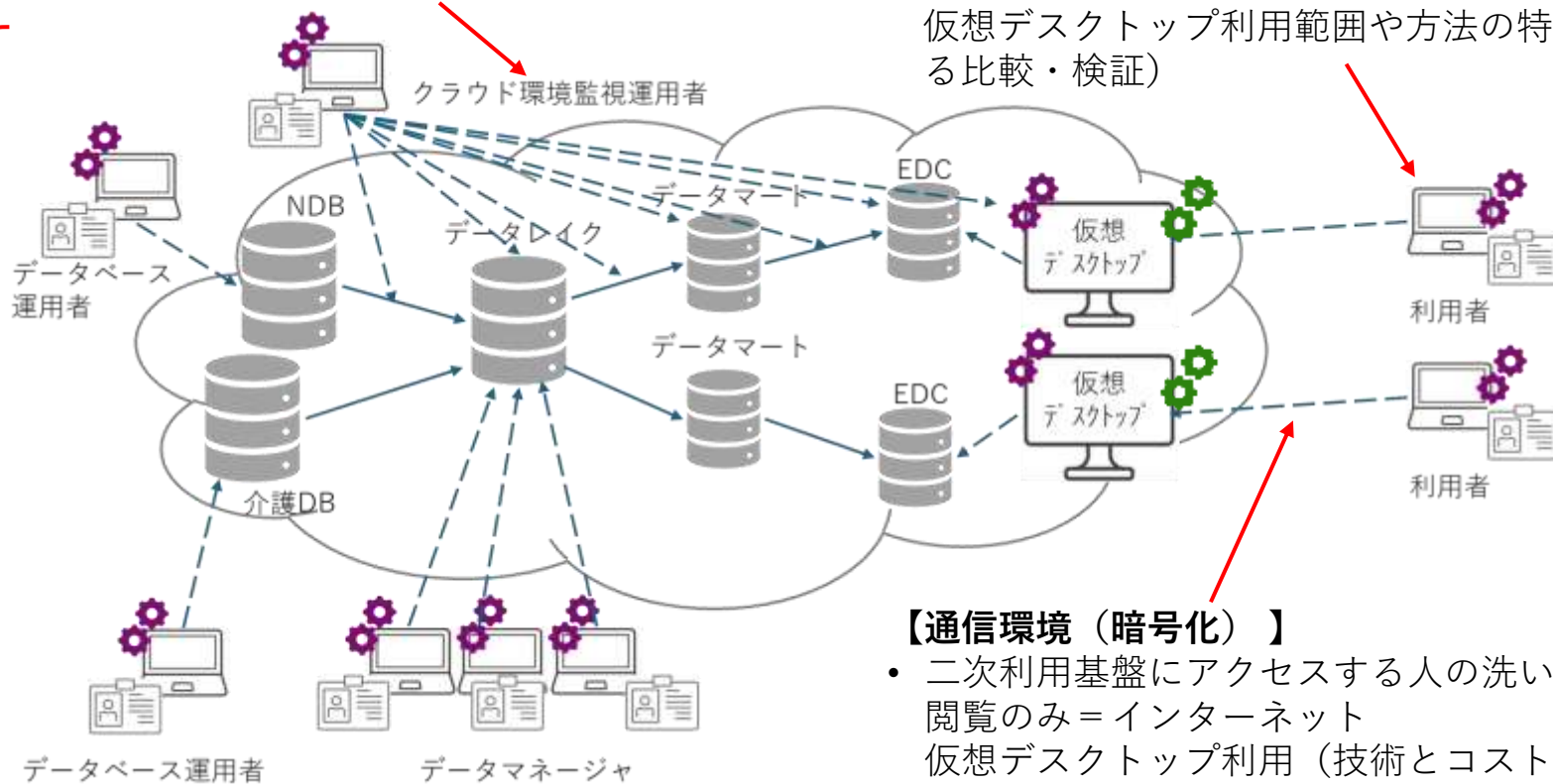
- ログの活用による予防と早期発見
- 極力自動化、人手の削減
- 分析ログの取捨選択と相関分析の実施（リアルタイム）
- 証跡として保管するものの管理（アーカイブ）

## 【安全管理措置】

- 技術的  
ゼロトラスト
- 物理的  
利用場所の制限（公共の場所はNGなど）  
責任範囲の明確化（利用環境、提供環境）
- 人的  
利用前研修  
誓約書の提出  
アクセス権限別の管理  
ルール周知徹底

## 【緊急対応】

- アラート発報ルール設定
- 緊急度に合わせた対応ルール設定
- 体制整備



## 【認証方法（二要素認証）】

- 二次利用基盤にアクセスする人の洗い出しと個別の設定情報閲覧のみ、仮想デスクトップ利用（利用方法の分類）等
- 2つの要素を用いた本人の認証+認可コントロールも必要  
仮想デスクトップ利用範囲や方法の特定（技術とコストによる比較・検証）

## 【通信環境（暗号化）】

- 二次利用基盤にアクセスする人の洗い出しと個別の設定閲覧のみ=インターネット  
仮想デスクトップ利用（技術とコストによる比較・検証）  
= オープンVPN (IPSec)  
※閉域網や暗号化の過信は禁物（外務省事案）

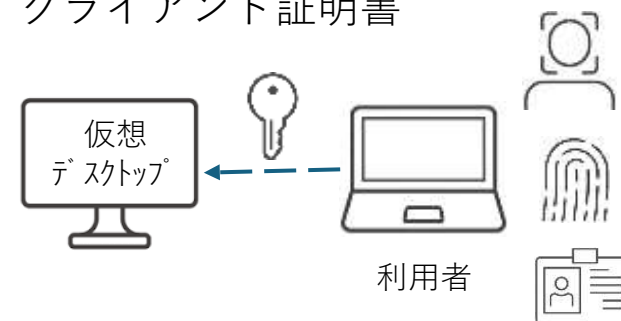
# 認証方式

## 【二要素認証 ふたつの認証方法の組合せ】

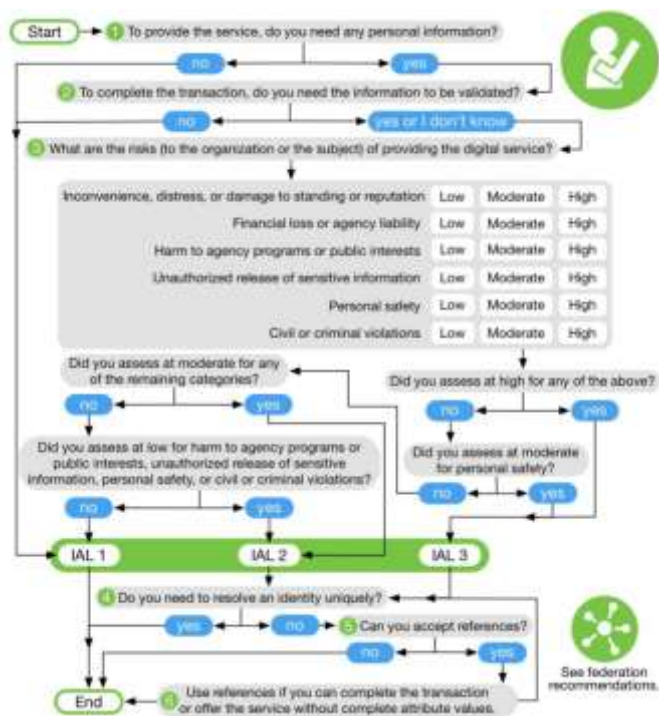
自分が知っているもの: パスワード、暗証番号、セキュリティ質問など

自分が持っているもの: 物理的なトークン、またはデジタルトークン (XXAuthenticator)、クライアント証明書

自分自身: “顔” や“指紋”、“静脈”のような生体認証



NISTSP800-63-3 電子認証に関するガイドライン (米国国立標準技術研究所)



### 本人確認 (IAL)

- Lv 1 : 自己申告による登録でよい
- Lv 2 : リモートまたは対面による確認
- Lv 3 : 対面での有資格者による確認

### 認証プロセス (AAL)

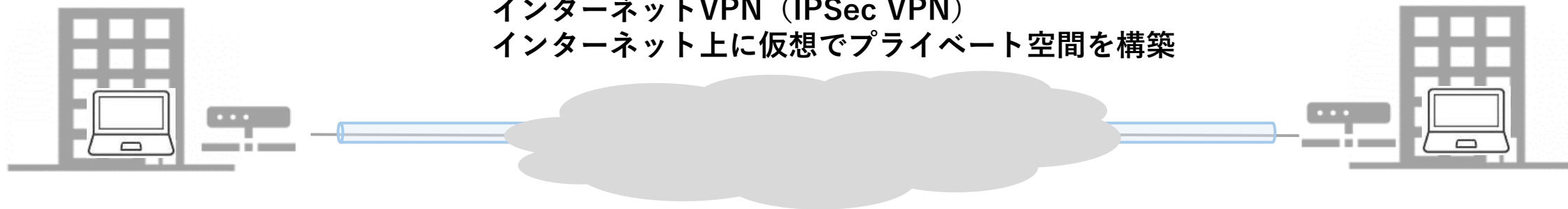
- Lv 1 : 単要素
- Lv 2 : 二要素必要、二要素目の認証手段はソフトウェアベースで可
- Lv 3 : 二要素必要、かつ二要素目はハードウェアを用いること

### 認証連携 (データ流通の仕組み) (FAL)

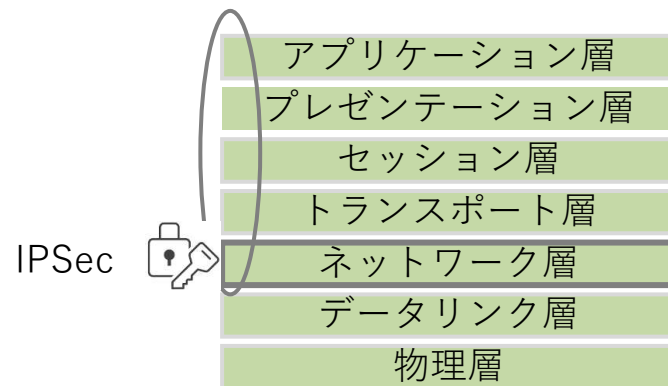
- Lv 1 : アサーションへの署名
- Lv 2 : アサーションへの署名に加えて対象のサービス提供者だけが複合できる暗号化
- Lv 3 : 上記二つに加えてさらに複雑なアサーションの利用

# 通信暗号化

インターネットVPN (IPSec VPN)  
インターネット上に仮想でプライベート空間を構築



通信階層と暗号化のポイント



<参考> IP-VPN (閉域網≠専用線)  
通信事業者が用意する閉じられた網



ルーターから利用者側の管理

# Visiting環境データ保護のルール例 英国Genomics England (GEL) TRE 環境



< “Airlock” を使用したファイルのインポートとエクスポート >  
 = 解析環境は、貸出図書館ではなく、閲覧図書館であるべき =  
 研究環境からエクスポートできるデータを制限するとともに、事前の承認を  
 必須とする。

アプリケーションの画面から、必要事項を入力し申請する。

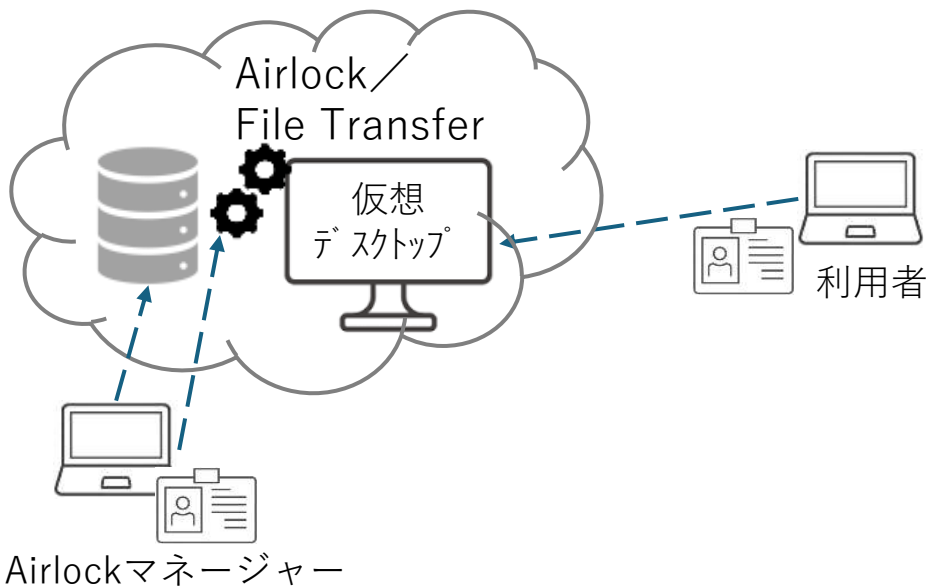
< 申請時の入力内容 >

- 研究チームまたは産業フォーラム等所属チームの番号
- 研究者登録ID
- プロジェクト番号
- エクスポートを希望するファイル（解析環境外で操作するデータや、持出し後公開する予定のもの全て）
- ファイルをエクスポートする理由とその意図（詳細に）

< インポートする際のルール（抜粋） >

- ウィルスチェック後のファイルであること（GEL側でもチェック）
- GEL内部でチェックし拒否の場合は理由と共に返送

意図的なセキュリティ侵害は、研究者及び所属する組織全体を利用停止に。  
 また個人を同定する行為は法律違反であり刑事告訴または高額の罰金刑。



# 安全管理措置

- 利用者環境

  - 技術的安全管理措置

    - ✓ データダウンロード、アップロードの申請制
    - ✓ ウィルス対策ソフトの導入とスキャンの実施
    - ✓ OSの最新化
    - ✓ 端末ロックの設定
    - ✓ フリーWifiの利用禁止 など

  - 物理的安全管理措置

    - ✓ 公共の場所（自組織内の外来者が入れるスペース含む）での利用禁止

  - 人的安全管理措置

    - ✓ 利用開始前のトレーニング受講義務化
    - ✓ 利用規程の周知徹底（罰則の規定を含む）
    - ✓ 申請手続きの標準化（理解促進と厳守（守りやすい環境整備））

- 提供者環境

  - ⇒各種ガイドラインに準じた対策

    - 医療情報安全管理ガイド、政府機関等の対策基準策定のためのガイドライン

  - ⇒対策実行のモニタリング（技術：SOC（専門的な監視） ルール：標準化と監査）

# ログの活用と緊急対応

**「置いただけ」「見ているだけ」にならないように**

- **システムの全体構成図／情報資産のリスト**
  - どこに何があり、何を監視しているか可視化する。情報資産の増減や配置変更などがあればアップデートする。
- **収集している（できる）ログの種類把握と発報ルール設定**
  - 個別のログを統合的に管理する。またログ分析結果に基づき対応ができるよう、分析ルールを作成する。（SIEM等の活用）
  - クラウド特有のログの管理方法がある点には十分配慮が必要。（専門知識を有する監視者による監視）
- **発報を受けた後の対応フロー設定**
  - 発報のレベルや内容によって、自動化し対応するものと、人を介して対応するものに分類し運用する。
  - どのレベルの発報があった場合エスカレーションするか、またエスカレーション先も明確にする。
- **緊急対応マニュアル・フローの作成**
  - CSIRT\*立上げ基準の設定
  - CSIRT立上げ後の対応手順やフローを設定する。複数の部門等の連携が必要になる点に留意が必要。



# セキュリティ対策の基本

- 日常的な監視を怠らない。
  - 常に状態を監視し、不審な挙動に早期に気づき、未然防止や被害の局所化が可能なように
- 緊急時の具体的な対策方法を規定する。
  - 問題発生時の緊急対応手順や体制 (CSIRT)の整備
  - 緊急対応（意思決定）が可能なログ収集と分析（ログは一定期間保管する）
  - セキュリティの要注意情報の活用と緊急時の遮断
- 対策費用と効果のバランスのとり方
  - 情報資産の場所や機微性に照らして、対策の配置を検討
  - 人の負荷を軽減するための方策（自動化、AI）
  - 利用者の啓発
- ITは守りの「道具」。利用者及び運用者双方の手順や規程・ルールが必要。
  - 利用者及び運用者の責任範囲とその内容の明確化

# 医療等情報にかかる民間データベースの 現状と連携上の課題

一般社団法人医療データベース協会 ディレクター  
情報処理安全確保支援士・弁護士・弁理士  
足立 昌聰

# 医療データベース協会（AMDJ）について

- 医療、介護及び健康等に関連し民間事業者によって運営されるデータベースが安心して利活用されるために、意見形成や環境づくりを行い、また、その活動を通して、**医療データベース業界を健全に発展**させ、患者を含む生活者のQOL向上へと貢献することを目的とし、2014年4月に設立。
- 2017年2月より、**認定個人情報保護団体**として、会員企業の個人情報の適正な取扱いを支援。**医療データベース事業者の”trust”（信任）の確保**のための活動を展開。

## “安心・安全な医療DBをすべての人に”

近年、日本における医療データの利活用への社会的要請が益々高まっております。しかし、国際的な疫学研究にも通用するデータ品質の保証、患者やデータ提供者のプライバシーに十分配慮したデータベース利活用のあり方など、医療・公共団体・企業などの関係各界を横断して解決されるべき課題が散見されます。

当協会では、各界の関係者が安心して医療DBを活用でき、ひいては医療DB産業が医療の質の向上や公衆衛生の発展に貢献できるための環境整備について、民間のデータベース実務家としての立場から意見形成を行ってまいります。

- 2024年3月1日現在、正会員9社、賛助会員2社の計11社が加盟。
- 日本薬剤疫学会が公表している「日本における臨床疫学・薬剤疫学に応用可能なデータベース\*」で紹介されている営利企業11社中8社が加盟。

\* [https://www.jspe.jp/mt-static/FileUpload/files/JSPE\\_DB\\_TF\\_J.pdf](https://www.jspe.jp/mt-static/FileUpload/files/JSPE_DB_TF_J.pdf)

# 民間の医療データベース

## 【データの由来による分類】

- 民間の医療データベースは、概ね、データ生成時の個人情報取扱事業者・行政機関等の種類に大別して、①**医療機関由来**、②**保険者由来**、③**保険薬局由来**の3種の系統が存在する。

## 【データの法的性質による分類】

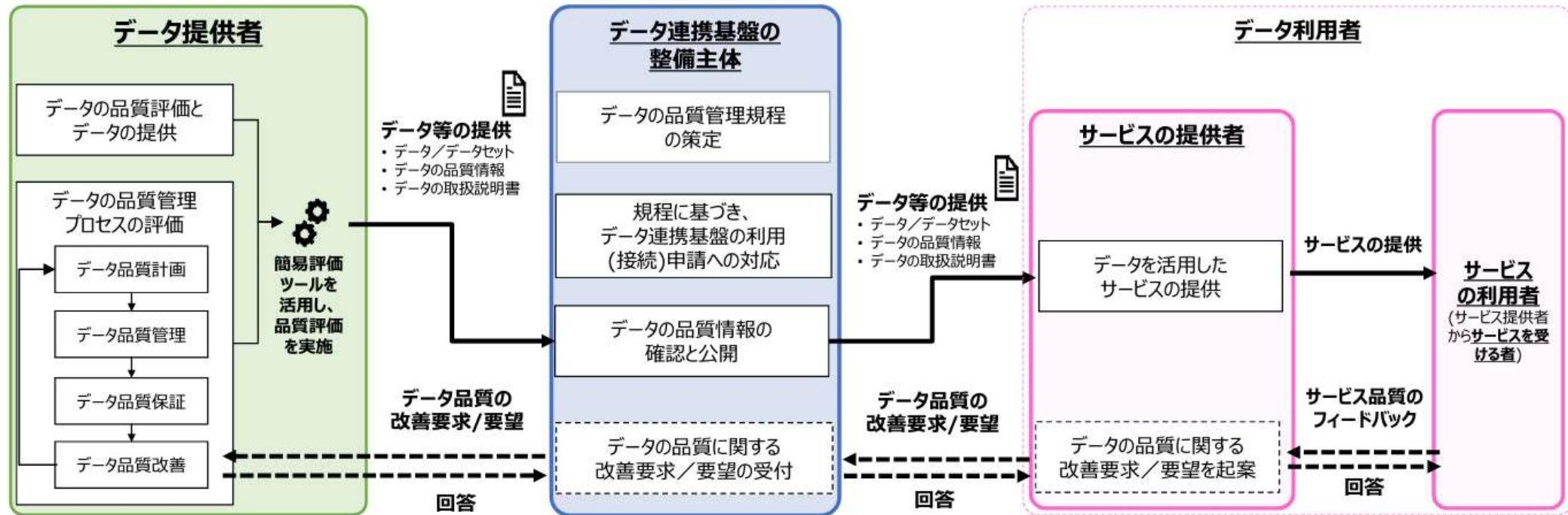
- 2024年3月現在、民間の大規模医療データベースはすべて、**個人情報保護法に基づく匿名加工情報（行政機関等匿名加工情報を含む。）**で構成されている。本人同意に基づく個人データの取得を介するケースは、PHR/CMP等を介した前向き臨床研究等など限定的。
- 2024年3月現在、次世代医療基盤法に基づく匿名加工医療情報の提供は認定された3スキームにとどまり、**営利企業が認定匿名加工医療情報作成事業者になった例はない。**

## 【データの属性による分類】

- 電子カルテ等の**アウトカム**（診療行為を実施した結果に関する情報）を含むものと、**レセプト等のインプット**（診療行為の実施に関する情報）しか含まないものに大別される。
- 次世代医療基盤法では、アウトカム（公的DBですでにカバーされているものを除く。）を含まないデータベースは認定されない。

# データの品質管理に対する考え方

- データの品質管理については、ISO/TS 8000で国際標準化されており、データ提供者からデータ利用者に至るプロセス\*について、品質管理のPDCAサイクルを回すことが適切。
- レギュラトリー領域などでは、プロセスに関連するデータ特性（追跡可能性等）も重要。



\* 2023.9内閣府地方創生推進事務局「データ連携基盤を通して提供されるデータの品質管理ガイドブック」より  
[https://www.chisou.go.jp/tiiki/kokusentoc/supercity/pdf/supercity\\_230926\\_guidebook\\_honsi.pdf](https://www.chisou.go.jp/tiiki/kokusentoc/supercity/pdf/supercity_230926_guidebook_honsi.pdf)

# データ自体の品質特性

- データ品質の国際標準規格であるISO/IEC 25012では、**15種の品質特性**が定義されている。
- 正確性、完全性及び一貫性（基礎的品質特性）のほか、**最新性や効率性などの商業的な付加価値となり得る品質特性（付加的品質特性）により市場競争が行われている。**

正確性 (Accuracy)	データに誤りがなく、信頼性がある状態か
完全性 (Completeness)	必要な項目が網羅されているか、データに誤りや欠落がなく完全な状態であるか
一貫性 (Consistency)	データに矛盾はないか、データ間で整合性が取れている状態か
信憑性 (Credibility)	出所や更新日が明示されているか
最新性 (Correctness)	更新期間や最終更新日が明示されているか
アクセシビリティ (Accessibility)	環境依存文字等が使用されていないことが明示されているか
標準適合性 (Compliance)	標準に準拠したデータ形式であることが明示されているか
機密性 (Confidentiality)	アクセス制限等の対策が講じられていることが明示されているか
効率性 (Efficiency)	効率化のため、コードの割り当て等を行っていることが明示されているか
精度 (Precision)	データの精度について明示されているか
追跡可能性 (Traceability)	データの変更履歴（変更者／変更日等）が明示されているか
理解性 (Understandability)	データ全体や項目ごとにメタデータが明示されているか
可用性 (Availability)	利用時間帯等について明示されているか
移植性 (Portability)	標準的な形式でデータをエクスポートできることが明示されているか
回復性 (Recoverability)	バックアップ等の障害時の対応方針が明示されているか

## 品質特性における協調領域

- **基礎的品質特性、アクセシビリティ及び追跡可能性**については、データの加工段階ではもはや**データ欠損が回復できないし、文字コード置換等の作業は付加価値となりづらく、上流の段階でこれらの品質特性が担保されることが強く期待される。**

<b>正確性 (Accuracy)</b>	データに誤りがなく、信頼性がある状態か
<b>完全性 (Completeness)</b>	必要な項目が網羅されているか、データに誤りや欠落がなく完全な状態であるか
<b>一貫性 (Consistency)</b>	データに矛盾はないか、データ間で整合性が取れている状態か
<b>信憑性 (Credibility)</b>	出所や更新日が明示されているか
<b>最新性 (Correctness)</b>	更新期間や最終更新日が明示されているか
<b>アクセシビリティ (Accessibility)</b>	環境依存文字等が使用されていないことが明示されているか
<b>標準適合性 (Compliance)</b>	標準に準拠したデータ形式であることが明示されているか
<b>機密性 (Confidentiality)</b>	アクセス制限等の対策が講じられていることが明示されているか
<b>効率性 (Efficiency)</b>	効率化のため、コードの割り当て等を行っていることが明示されているか
<b>精度 (Precision)</b>	データの精度について明示されているか
<b>追跡可能性 (Traceability)</b>	データの変更履歴（変更者／変更日等）が明示されているか
<b>理解性 (Understandability)</b>	データ全体や項目ごとにメタデータが明示されているか
<b>可用性 (Availability)</b>	利用時間帯等について明示されているか
<b>移植性 (Portability)</b>	標準的な形式でデータをエクスポートできることが明示されているか
<b>回復性 (Recoverability)</b>	バックアップ等の障害時の対応方針が明示されているか

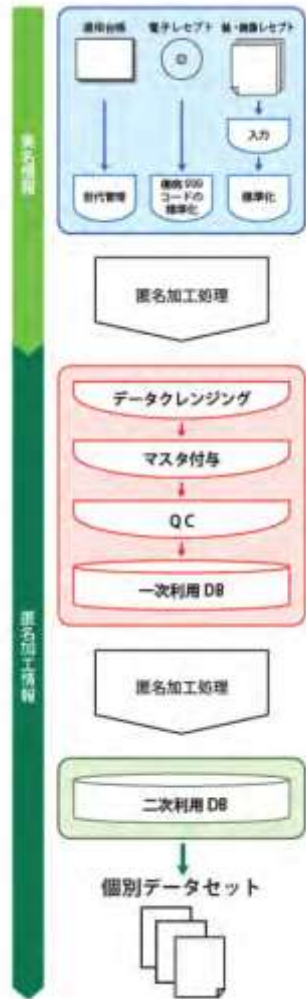
## 品質特性における競争領域

- 効率性や可用性といった付加的な品質特性は求められる水準が需用者により異なり、基礎的品質特性を超えた「使いやすさ」は競争領域であり、プロセスも複雑になりやすい。
- 標準適合性は、需用者により適合すべき標準の範囲が異なる。

正確性 (Accuracy)	データに誤りがなく、信頼性がある状態か
完全性 (Completeness)	必要な項目が網羅されているか、データに誤りや欠落がなく完全な状態であるか
一貫性 (Consistency)	データに矛盾はないか、データ間で整合性が取れている状態か
信憑性 (Credibility)	出所や更新日が明示されているか
最新性 (Correctness)	更新期間や最終更新日が明示されているか
アクセシビリティ (Accessibility)	環境依存文字等が使用されていないことが明示されているか
標準適合性 (Compliance)	標準に準拠したデータ形式であることが明示されているか
機密性 (Confidentiality)	アクセス制限等の対策が講じられていることが明示されているか
効率性 (Efficiency)	効率化のため、コードの割り当て等を行っていることが明示されているか
精度 (Precision)	データの精度について明示されているか
追跡可能性 (Traceability)	データの変更履歴（変更者／変更日等）が明示されているか
理解性 (Understandability)	データ全体や項目ごとにメタデータが明示されているか
可用性 (Availability)	利用時間帯等について明示されているか
移植性 (Portability)	標準的な形式でデータをエクスポートできることが明示されているか
回復性 (Recoverability)	バックアップ等の障害時の対応方針が明示されているか



# クレンジング、コード化、標準化等の実際例（一部）

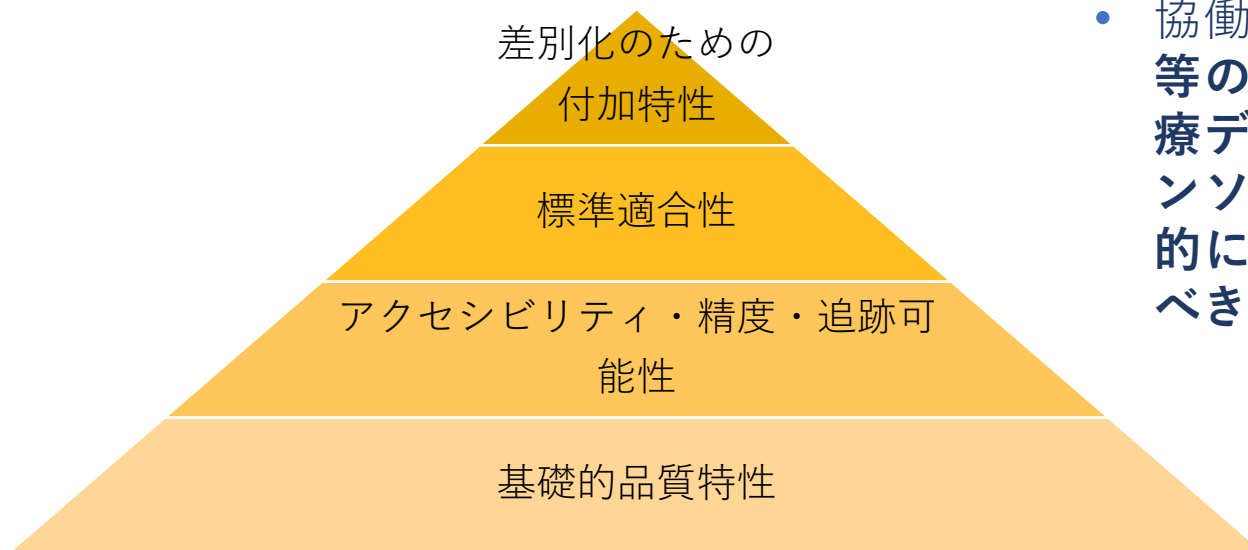


- 取込データの最初のバリデーション（基礎的品質特性、精度の確認等）やクレンジングでは、検査値、投与単位等の学会公式基準等への変換や、いわゆる「その他のX」系データのコード化などを行う。
- **コードマッチしない項目のコード化は、各社が蓄積した辞書や自然言語処理に基づいて行われている（競争領域）。**
- 標準適合性のためのマスタのメンテナンスは、各マスタについて月1回以上であり、**診療報酬情報提供サービスや薬価基準収載医薬品等の公的な情報については、改定の都度マスタを更新し、過去分との統合的な解析を確保している。**
- 医薬品毎の添付文書に含まれる情報等の付加や解析ニーズの高いコードの付加、多言語化等は、個別の需用者のニーズを踏まえ、匿名加工の前後で実施される。
- 全行程の工程数としては、3,000ステップを超えている\* ところもある。

\* 構成員個人の所属企業のユースケースに基づく

## 品質確保における官民協調と標準化の考え方\*

- 上流でなければ担保できない特性（基礎的品質特性、アクセシビリティ等）は官民で課題感を共有し、商用・学術利用を問わずに求められる標準化を進めるべき。
- 標準適合性のためにマスタのメンテナンス等で民間事業者が相当の資本を投下している知的財産を官民の連携プラットフォームのために利用する必要がある場合は、標準規格必須特許（SEP）におけるFRAND\*の考え方を適用し、官民構成員の当該マスタの利用機会の確保と同時に、相当の利益回収の機会が確保されるべき。



- 協働領域・競争領域を含め、学術、製薬企業等の医療等データの二次利用者に加えて、医療データベース事業者を交えた連携基盤（コンソーシアム等）で、品質確保のための商業的に持続可能な仕組みの創出を官民で目指すべき。

\*FRAND (Fair, Reasonable, And Non-Discriminatory terms and conditions): ライセンスに際しての金額や条件が公正、合理的かつ非差別的であること

\* 構成員個人の私見であり、AMDJ加盟企業の意見を代表するものではない。

## 二次利用基盤全体に関する技術的論点

### 【データプライバシーに関する論点】

- 医療等データによる**個人識別性の問題（データプライバシー）**と、**本人の社会的実在の特定（いわゆる身バレ）の問題**を区別して議論する必要がある。
- k匿名性等の統計的な匿名性（データプライバシー）を議論する場合は、**攻撃者の具体的な想定を置かないまま議論をすると統計的な定性的・定量的な評価手法の選択ができない。**
- **標準適合性の確保の前後で匿名性が変わる（同値のsubject数が変わる）**ため、**どの段階で仮名化・匿名化を実施するのか**を議論する必要がある。解析環境に解析者が持ち込んだマスタでも匿名性は変化することにも留意する必要がある。
- **アドホックな解析条件で匿名加工の粒度を可変させるのか、定期的に静的な匿名加工情報のデータベースを構築するのか（前者の方が幅広のニーズに対応するが、設計が難しい）。**

### 【本人関与に関する論点】

- 母集団への取込へ本人関与について、オプトイン・オプトアウトで技術的な難しさには大差がない。それよりも、**本人関与の機会の有無のフラグ管理をネットワーク全体で確保する必要**があり、次世代医療基盤法のインシデントもここに問題が集中している。
- 死者の情報は個人情報ではないが、生前に本人関与の機会がないと取り込めない制度設計の場合は、**随時死亡情報が連携されない限り、取込可否の判定ができない**（本人関与を必要とする理念的背景が技術的設計に影響する。）。

# 民間の医療データベースとの連携に関する論点

## 【匿名加工情報との連携の法的スキーム】

- 次世代医療基盤法上の仮名加工医療情報と異なり、個人情報保護法上の匿名加工情報（行政機関等匿名加工情報を含む。）は、**匿名加工情報取扱事業者に提供された時点で、個人識別可能な情報を含まない。**
- **連結可能とするために被保険者番号を保持した時点で、その提供は個人データ・保有個人情報の第三者提供**に当たり、連結は新たな個人情報の取得となるため、個人情報取扱事業者との間で連結可能な状態での提供を行うことは（特別法があっても）困難。

## 【連携基盤上での個人データの取扱主体】

- 民間DBと連結したデータの解析者は現行法上は仮名加工情報の取扱いの主体となるため、安全管理措置を確保すべき主体となる。**安全管理措置の内容と責任分界点を適切に設定される**必要がある。

## 【商用データベースのライセンス管理】

- 連携された民間データベース由来のデータはHIC等の解析基盤上で解析されることとなると思われるが、その場合に**解析者の実際の利用形態**（解析者固有のデータとして解析環境への持ち込みを行いのか、そもそもカタログ収載されるのか等）により、データの出力手法が異なる。

Keywords: 医療等情報, 医療データベース協会, AMDJ, 匿名加工情報, 仮名加工情報, 匿名加工医療情報, 仮名加工医療情報, 行政機関等匿名加工情報, 標準化, 基礎的品質特性, 付加的品質特性, データプライバシー

## AMDJ 一般社団法人 医療データベース協会

ディレクター  
情報処理安全確保支援士・弁護士・弁理士  
足立 昌聰

本資料に関するご質問等: [madachi\[アット\]jmdc.co.jp](mailto:madachi@jmdc.co.jp)

2024.3 医療等情報の二次利用に関するワーキンググループ 技術作業部会資料